VARIATIONS IN CARBON MONOXIDE, NITRIC OXIDE, AND DETOXIFICATION GENES, INTERACTIONS WITH MATERNAL SMOKING, AND ASSOCIATIONS WITH PREECLAMPSIA: A MOTHER-CHILD DYAD ANALYSIS

Anna E. Bauer

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology in the Gillings School of Global Public Health.

Chapel Hill
2017

Approved by:

Stephanie M. Engel

Christy L. Avery

Tracy A. Manuck

Andrew F. Olshan

Clarice R. Weinberg

ABSTRACT

Anna E. Bauer: Variations in carbon monoxide, nitric oxide, and detoxification genes, interactions with maternal smoking, and associations with preeclampsia: a mother-child dyad analysis
(Under the direction of Stephanie M. Engel)

Preeclampsia is a serious pregnancy complication with limited treatment. Etiology is hypothesized to originate during placentation, and may have both maternal and fetal contributions. There is a well-established enigmatic inverse relationship between maternal smoking and preeclampsia. A plausible biological explanation for this relationship is through response to cigarette smoke components, via vasodilation or activation of smoking detoxification pathways. Examining genes in these pathways and their modification by smoking, while incorporating maternal and child genetic contributions, could provide support for a genetic or related biological mechanism.

We conducted a nested case-control study within the Norwegian Mother and Child Birth Cohort of 1,545 case-pairs and 995 control-pairs from 2,540 validated dyads (2,011 complete pairs, 529 missing mother or child genotype). For aim 1, we selected 1,518 single nucleotide polymorphisms (SNPs) in nitric oxide and carbon monoxide signaling pathways. For aim 2, we analyzed these and 397 additional SNPs in smoking detoxification pathways for their modification by maternal smoking during placentation. We used log-linear Poisson regression models and likelihood ratio tests to assess maternal and child effects and included a SNP by smoking interaction term to assess maternal and child genotype-smoking interactions.

The child variant, rs12547243 in adenylate cyclase 8 (*ADCY8*), was associated with an increased risk (RR=1.42 [95% CI: 1.20, 1.69] for *AG* vs *GG*, RR=2.14 [1.47, 3.11] for *AA* vs *GG*,

Q=0.03). We also found suggestive associations of SNPs in *PDE1C* for preeclampsia sub-phenotypes. We found limited evidence for multiplicative SNP by smoking interaction after correction for multiple comparisons.

This study uses a novel approach to disentangle maternal and child genotypic effects of smoking-related genes on preeclampsia. Our findings do not provide strong support that the inverse smoking-preeclampsia relationship is due to a genetic effect in these pathways, although our power was limited due to the low prevalence of smoking in this population. Dyad methods and gene-environment interaction analysis may be useful for the study of pregnancy outcomes, particularly preeclampsia. Larger populations, such as multi-cohort consortia combined with these evolving methods may be necessary to dissect this enigmatic association.

ACKNOWLEDGEMENTS

I would like to thank my dissertation committee, Drs. Stephanie Engel, Christy Avery, Tracy Manuck, Andy Olshan, and Clarice Weinberg for their guidance, expertise, and support on this project. I would particularly like to thank my chair, Stephanie Engel, for her invaluable mentorship, as both a mentor of epidemiology and of life. I would also like to gratefully acknowledge the contributions of the following people: Drs. Min Shi and Angela Mazul for their extensive statistical genetics support; Drs. Quaker Harmon, Michael Wu, and Jenny Yang, for their knowledge and assistance of the project data management and study design; Dr. Jason Luo for genotyping of the samples; Drs. Kari Klungsøyr, Lill Trogstad, and Per Magnus for their support from the MoBa Study and Norwegian Institute of Public Health; and Drs. Nicholas Williams, Ralph McGinnis, and Linda Morgan for their assistance with the replication study.

I am extremely grateful for the contributions of the women and families who participated in the Norwegian Mother and Child Cohort Study.

I would like to acknowledge the financial support provided by a pre-doctoral training grant from the Eunice Kennedy Shriver National Institute for Child Health and Human Development and support for the study from the National Institute of Environmental Health Sciences, the Norwegian Ministry of Health, and the Norwegian Research Council.

I am grateful for all of my peers and colleagues at UNC including those in the student services office, professors, postdocs, and students who have provided support during my time here. I would particularly like to acknowledge Stephanie Engel's advisee group, as well as my writing group: Chris Gray, Melanie Napier, Sabrina Zadrozny, and Dana Pasquale.

Finally, I'd like to acknowledge my family. I thank my partner Chris Paul, for his never-ending patience, encouragement, and a shared joy in the excitement of research; my parents and in-laws, Gary and Sheryl Bauer and John and Jill Paul for their continuous support in more ways than can be articulated; and my daughters, Emma and Lucy, for the daily reminder of why it is I study maternal and child public health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACOG	American College of Obstetricians and Gynecologists
AhR	Aryl hydrocarbon receptor
BMI	Body mass index
cAMP	Cyclic adenosine monophosphate
CEPH	Centre d'etude du polymorphisme humain
cGMP	Cyclic guanosine monophosphate
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
CI	Confidence interval
CO	Carbon monoxide
CYP	Cytochrome P450
DAG	Directed Acyclic Graph
EM	Expectation maximization
eNOS	Endothelial nitric oxide synthase
FDR	False Discovery Rate
GEWIS	Genome-Environment Wide Interaction Study
GOPEC	Genetics of Preeclampsia consortium
GST	Glutathione-S-transferase
GWAS	Genome-Wide Association Study
GxE	Gene by environment
HELLP	Hemolysis, elevated liver enzymes, low platelet count syndrome
HIF1A	Hypoxia inducible factor 1-alpha
HO	Heme oxygenase
HWE	Hardy Weinberg Equilibrium
IBD	Identity by descent

ICD-10	International Classification of Diseases, 10th edition
LD	Linkage disequilibrium
LEM	Likelihood expectation maximization (software)
LOD	Logarithm of the odds score
MAF	Minor allele frequency
MBRN	Medical Birth Registry of Norway
MoBa	Norwegian Mother and Child Cohort Study
NO	Nitric oxide
NOS	Reactive nitrogen species
PE	Preeclampsia
PCA	Principal Components Analysis
PlGF	Placental growth factor
QC	Quality control
ROS	Reactive oxygen species
RR	Risk ratio
sEng	Soluble endoglin
sFlt-1	Soluble fms-like tyrosine kinase-1
SGA	Small for gestational age
SNP	Single nucleotide polymorphism
VEGF	Vascular endothelial growth factor
WTCCC2	Wellcome Trust Case Control Consortium 2

CHAPTER I. INTRODUCTION AND SPECIFIC AIMS

A. Introduction

Preeclampsia, characterized by gestational hypertension and proteinuria, occurs in 2-8% of pregnancies.[1] It is a serious complication of pregnancy, and resulting sequelae can be harmful to both the mother and child. Preeclampsia can progress to eclampsia, which is defined by seizures and is a potentially life threatening condition.[2] Low-dose aspirin may help prevent preeclampsia in high-risk women,[3] but treatment is often limited to delivery,[4] which may pose serious risks to the infant if preterm.

Preeclampsia is of unknown etiology, but is hypothesized in part to originate from processes occurring during formation of the placenta.[2,4] During normal placentation, fetal cytotrophoblasts invade the maternal decidua and penetrate the walls of adjacent maternal spiral arteries. Here, the arteries lose smooth muscle, greatly increasing vascular dilation. This process does not completely occur with preeclampsia, reducing vascular capacity.[5]

There is a well-established enigmatic inverse association with smoking and preeclampsia.[6] One hypothesis for this association is through the powerful, vasodilatory action of carbon monoxide (CO) and nitric oxide (NO),[7,8] which are produced by combustion of cigarettes[9,10] as well as endogenously in the body.[11–16] We aimed to determine if genetic variants in pathways related to CO and NO signaling are associated with preeclampsia, which may occur through changes in endogenous production of these compounds. Three canonical pathways are particularly relevant to CO and NO activity: 1) endothelial nitric oxide synthase (eNOS) signaling, 2) heme degradation, and 3) hypoxia-inducible factor 1-alpha (HIF1A)

signaling. The eNOS signaling pathway primarily describes the endogenous synthesis of NO from L-arginine, heme degradation describes the breakdown of hemoglobin into CO and bilirubin, and HIF1A signaling describes regulation of oxygen homeostasis and response to hypoxia important for both CO and NO activity. Genetic polymorphisms in all of these pathways exist,[17,18] and several are associated with cardiovascular diseases that share features with preeclampsia.[12,17,19–21] Inconsistent associations with preeclampsia have been found for three genes in these pathways (*HMOX1, eNOS, iNOS*).[22–24] However, sample sizes in these studies were small, however, and none investigated entire pathways, gene by environment interactions, or the combination of maternal and child DNA.

We also aim to determine if associations of SNPs involved in response to smoking components with preeclampsia are modified by an exogenous exposure (maternal smoking). For this aim, we included additional smoking detoxification genes, particularly those encoding the CYP and GST enzymes and involved in the pathways: 1) xenobiotic metabolism 2) aryl hydrocarbon receptor signaling, 3) glutathione-mediated detoxification, 4) nicotine degradation II and, 5) nicotine degradation III.

This study is a nested case-control study ancillary to the Norwegian Mother and Child Cohort Study (Den norske mor og barn-undersøkelsen (MoBa)). MoBa is a longitudinal study following women and their children from pregnancy through childhood.

The study uses a mother-child dyad design including 1564 validated preeclampsia cases and 999 validated non-preeclampsia controls, and 2563 total mother-child pairs. Because the placenta is of fetal origin and because mothers and their offspring share half of their genetic material, fetal genetics may play a role in the development of preeclampsia. Dyad designs are most appropriate in circumstances like this, where both maternal and child genetics have independent associations with the outcome, but the association of child genetics is inherently confounded by maternal genetics. Disentangling these features allows us to ascertain whether associations are primarily due to a maternal or fetal genetic component.

B. Specific Aims

Specific Aim 1

Determine the association of genetic variants within genes in nitric oxide and carbon monoxide signaling pathways with risk of preeclampsia. Pathways include endothelial nitric oxide synthase signaling (eNOS), heme degradation, and hypoxia-inducible factors 1-alpha (HIF1A).

Aim 1a: Evaluate the association between maternal and child single nucleotide polymorphisms in these pathways and risk of preeclampsia.

Aim 1b: Evaluate the associations between variants in these pathways on risk of preeclampsia stratified by preeclampsia subtype: early-onset preeclampsia, severe preeclampsia, preeclampsia with delivery <34 weeks, and preeclampsia accompanied by small for gestational age (SGA).

Specific Aim 2

Assess gene by environment interaction of maternal smoking and genetic variants in carbon monoxide and nitric oxide signaling pathways, as well as smoking detoxification pathways in their association with preeclampsia.

Aim 2a: Determine if maternal active smoking at the time of spiral artery remodeling modifies the effects of carbon monoxide and nitric oxide genes described in Aim 1.

Aim 2b: Determine if maternal active smoking at the time of spiral artery remodeling modifies the effects of other smoking detoxification genes. Pathways include xenobiotic metabolism, aryl hydrocarbon receptor signaling, glutathione-mediated detoxification, and nicotine degradation.

Hypotheses

Hypothesis 1: Variants in the eNOS, heme degradation, and HIF1A signaling pathways are associated with a change in risk of preeclampsia; direction of change varies by SNP function.

Hypothesis 2: These effects may have both a maternal and fetal component.

Hypothesis 3: Stronger associations may also be found in more severe preeclampsia phenotypes, such as early-onset preeclampsia (<34 weeks' gestation), severe preeclampsia, preeclampsia with early delivery, or preeclampsia complicated by SGA.

Hypothesis 4: We expect differential effects of some of the variant associations in carbon monoxide, nitric oxide, and smoking detoxification pathways among mothers who smoked during pregnancy as compared to those who did not.

Hypothesis 5: We further expect stronger smoking by gene interactions among mothers who smoked throughout the entire pregnancy rather than only during early pregnancy.

Currently, there are few effective therapeutic interventions for the treatment of preeclampsia,[2,4] though low-dose aspirin supplementation has been recommended for prevention in high-risk women.[25] A better understanding of the biological processes leading to or protecting from preeclampsia may reveal new therapeutic targets, with the goal of reducing maternal and infant morbidity and mortality.

CHAPTER II. REVIEW OF THE LITERATURE

A. Preeclampsia

The symptoms of preeclampsia and eclampsia have been recognized for ages, since the time of Hippocrates.[26] These symptoms later became recognized as eclamptic hypertension and named preeclampsia in 1897, and in 1906, magnesium sulfate was used as medical management for preeclampsia and eclampsia.[26] Despite the long-existing knowledge of this condition, treatment has changed very little in the last century; magnesium sulfate is still used for management of severe preeclampsia/eclampsia during and after labor, and low-dose aspirin use for high-risk women is the only widespread preventive intervention that has been shown to be successful.[3,25,27] Although preeclampsia is well-studied, it remains perplexing and not well-understood.

Definition

Preeclampsia is characterized by de novo gestational hypertension and proteinuria after 20 weeks of pregnancy. In this proposed study, preeclampsia will be defined by the 2002 American College of Obstetricians and Gynecologists (ACOG) diagnostic criteria,[27] which include:

1) Systolic blood pressure of at least 140 mm Hg or diastolic blood pressure of at least 90 mm Hg, occurring after 20 completed weeks of gestation in a woman with previously normal blood pressure, and
2) Proteinuria, defined as at least +1 on urine dipstick measurement.

There is substantial heterogeneity in the severity, clinical characteristics, and timing of symptom onset in preeclampsia. Severe preeclampsia meets the general criteria described above and at least one of the following additional requirements[27]:

1) Systolic blood pressure of at least 160 mm Hg or diastolic blood pressure of 110 mm Hg on at least two occasions, measured 6 hours apart and while the patient is at rest,
2) Urinary protein excretion of at least 5g/day in a 24-hr urine specimen or at least 3+ urine dipstick protein measurement on two samples collected 4 hours apart,
3) Oliguria or less than 500 ml of urine in 24 hours,
4) Cerebral or visual disturbances,
5) Pulmonary edema or cyanosis,
6) Epigastric or right upper quadrant pain,
7) Impaired liver function,
8) Thrombocytopenia, or
9) Fetal growth restriction, which will be defined as estimated fetal weight or birthweight below the 10th percentile at a given gestational week in our study, however, can also include femur length and blood flow as determined by ultrasound.[28]

Early-onset preeclampsia is considered that which is diagnosed prior to 34 weeks' gestation.

The ACOG Task Force on Hypertension in Pregnancy recommended new diagnostic guidelines in November 2013.[25] Of greatest importance, the task force eliminated the requirement of proteinuria as a diagnostic criterion due to the "syndromic nature" of preeclampsia. In the absence of proteinuria, preeclampsia is diagnosed as hypertension combined with 1) thrombocytopenia, 2) impaired liver function, 3) new development of renal insufficiency, 4) pulmonary edema, or 5) new-onset cerebral or visual disturbances. These indications are considered severe features of preeclampsia, however, high urinary protein and fetal growth restriction have been removed as indications of severe preeclampsia.[25] The older

guidelines were used as they were in place at the time of enrollment in the MoBa cohort and are the guidelines that have been most widely used in other studies. However, it is necessary to consider definitional changes while comparing associations with preeclampsia across studies. In recent years, multiple changes in disease criteria is a significant challenge to the synthesis of the etiological literature.

Burden

Preeclampsia occurs in 2-8% of pregnancies worldwide.[1] Incidence of preeclampsia has increased in the US[29] and in Norway in recent decades.[30,31] It is one of the most common serious pregnancy complications. In the mother, preeclampsia can quickly become severe and lead to eclampsia which can cause placental abruption, cerebral hemorrhage, hepatic rupture, renal failure, pulmonary edema, seizure, brain damage, and death.[2,27,32–34] It is one of the most common contributors of maternal morbidity and mortality worldwide.[2] Approximately 10-15% of maternal deaths in both industrialized and developing countries are attributed to preeclampsia and eclampsia, however, most of the maternal mortality due to these conditions occurs in low-income countries.[35,36] Women with preeclampsia are at risk for future poor health outcomes including preeclampsia in a subsequent pregnancy,[37,38] diabetes,[39] and cardiovascular disease.[40,41] Preeclampsia can also lead to poor pregnancy outcomes that can affect the fetus, such as reduced amniotic fluid and abnormal oxygenation and poor fetal outcomes such as encephalopathy, fetal growth restriction, preterm birth, and perinatal death.[2,42,43] Preeclampsia is also a common reason for medically-indicated preterm delivery.[5] In the US, approximately 42% of medically-indicated preterm births and 15% of all preterm births may be attributed to preeclampsia.[44] Long-term consequences of preeclampsia in the child can include later cardiovascular disease, hypertension, stroke, and venous thromboembolism[40] as well as outcomes associated with preterm birth, such as visual, motor, and cognitive deficits.[45]

Despite advances in perinatal care and extensive research on causes of preeclampsia, the incidence of this serious condition has not declined,[34,46,47] and prediction and prevention of the disease has not improved.[46–48] Salt restriction, zinc, magnesium, fish oil, vitamin C and E supplementation, diuretics, calcium supplementation, and aspirin have been studied as potential preventive strategies;[46,49,50] low-dose aspirin is suggested to be the most promising,[3] and it has been one of the few therapies now recommended for women at risk of preeclampsia.[2,25,46] The only definitive treatment of preeclampsia is delivery of the fetus to prevent development of maternal or fetal complications.[43] The use of antihypertensive drugs to control elevated blood pressure has not been found to alter the course of the disease,[43] however, may reduce the occurrence of comorbid pregnancy complications such as maternal stroke and renal complications. Throughout labor and delivery, labetalol, hydralazine, and nifedipine can be given to control severe hypertension[43] and magnesium sulfate can be administered for the prevention of seizures,[51–54] but these therapies only aim to reduce the most serious outcomes and do nothing to prevent incidence of the disease.[43]

Etiology

Although the etiology of preeclampsia is unknown, the most widely accepted hypothesis is that it originates in part through incomplete or ineffective placentation.[5] During normal placentation, cytotrophoblasts invade the decidua, moving into the adjacent spiral arteries and the myometrium. As they penetrate the walls of the spiral arteries, the arteries are remodeled, losing smooth muscle to become dilated and allow for more blood flow.[5] Animal studies show that in preeclampsia, this process is incomplete and leads to a cascade of maternal inflammatory responses, resulting in placental ischemia and hypoxia.[55–57] The hypothesis is that this ischemia and hypoxia then induces the release of soluble fms-like tyrosine kinase-1 (sFlt-1) and soluble endoglin (sEng) into maternal circulation,[58,59] inhibiting production of vascular endothelial growth factor (VEGF) and placental growth factor (PlGF).[57,60,61] Reduction in these

proangiogenic factors may lead to an antiangiogenic state and cause widespread maternal endothelial dysfunction[55,56] (See Figure 1)

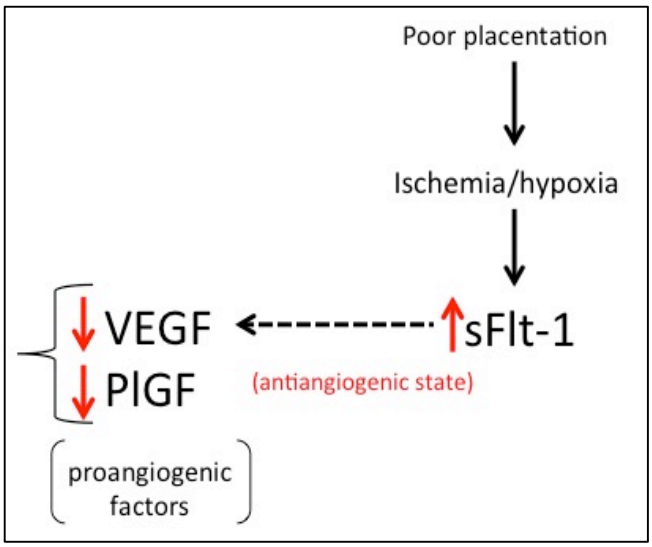


Figure 1. Antiangiogenic state resulting from poor placentation

In a longitudinal study of calcium supplementation, in which serum sFlt-1 and PlGF were measured over the course of pregnancy, sFlt-1 increased and PlGF decreased during the last two months of pregnancy even among normotensive controls, however, changes occurred earlier and were of greater magnitude in women who were later diagnosed with preeclampsia.[60] Additionally, these alterations were even greater in women with earlier onset preeclampsia (At 21 to 32 weeks: sFlt-1 1672 pg/ml with early preeclampsia vs. 935 pg/ml with term preeclampsia, $p<0.001$; PlGF 297 vs. 676 pg/ml, $p<0.001$. At 33 to 41 weeks: sFlt-1 8150 vs 2467 pg/ml, $p<0.001$; PlGF 73 vs. 370 pg/ml, $p=0.05$).[60] In a similar analysis of serum levels of 7,519 women of the Generation R study during early pregnancy (<18 weeks' gestation) and mid-pregnancy (18-25 weeks' gestation), low PlGF was associated with uterine artery resistance at both time points ($\beta= 0.033$, 95% CI:0.022-0.044 in early pregnancy and $\beta= 0.021$, 95% CI:0.010-0.032 in mid-pregnancy).[61] In women who were diagnosed with preeclampsia, sFlt-1 increased in the same time period (Delta sFlt-1 -0.032 ng/ml in controls vs 0.011 ng/ml in

cases, $p<0.05$) and there was a lower positive change in PlGF (21.4 pg/ml vs 14.7 pg/ml, $p<0.05$).[61]

It is speculated that preeclampsia that presents earlier in pregnancy (<34 weeks' gestation) and more severe forms of preeclampsia may be considered a separate phenotype and may have a different underlying etiology. Risk and protective factors differ for early-onset and late-onset preeclampsia. African-American race, chronic hypertension, and congenital anomalies are more strongly associated with early-onset preeclampsia, while younger maternal age, nulliparity, and diabetes mellitus are more strongly associated with late-onset preeclampsia.[62] In a study of differences in hemodynamic state, Valensise et al. found that patients with early preeclampsia had significantly more (15% versus 60%) bilateral notching of the uterine artery at 24 weeks than those with late onset,[63] which illustrates that there are potentially differing vascular blood flow patterns with timing of preeclampsia onset. Vascular resistance was also increased and cardiac output decreased in early preeclampsia, while prepregnancy BMI was higher in late preeclampsia than early-onset.[63] In a microarray study of gene expression within placental tissues, Nishizawa et al. identified eleven genes that were differentially up- or down- regulated in early-onset preeclampsia compared to late-onset preeclampsia.[64] Most studies differentiate subtypes of preeclampsia by time of diagnosis, however, it is important to note that timing is not independent of severity. Early-diagnosed cases are typically the most severe cases of preeclampsia.[65,66]

Risk Factors

Known risk factors for preeclampsia include nulliparity, previous preeclampsia, family history of preeclampsia, multifetal pregnancy, history of subfertility, chronic hypertension, diabetes mellitus, autoimmune disorders, kidney disease, obesity (high prepregnancy BMI), African-American race, and maternal age.[42,67,68]

Both young and old maternal age are risk factors for preeclampsia as is typical for most pregnancy-related morbidities, though some have found no difference in risk across age groups.[69,70] Other factors related to the pregnancy, such as parity, interpregnancy interval, and history of preeclampsia in a previous pregnancy are much stronger risk factors. First pregnancies have more than twice the risk of preeclampsia than later pregnancies.[68] Recurrence risk of preeclampsia is also high, and women who had preeclampsia in their first pregnancy have two to three times the risk in the general population of pregnant women.[68]

Traditional risk factors for cardiovascular disease and other comorbidities such as high pre-pregnancy body mass index, diabetes, chronic hypertension, and autoimmune disorders are also risk factors for preeclampsia, however, the relationship of these risk factors to preeclampsia is poorly understood.[1] Even when combined with pregnancy risk factors such as age and parity, it is difficult to predict who will be become preeclamptic.[4]

B. Relationship of Smoking and Cigarette Smoke Components

Epidemiology of Smoking and Preeclampsia

One of the most consistent associations with preeclampsia is maternal smoking (Figure 2). Maternal smoking is associated with as much as a 50% reduced risk of preeclampsia despite increasing risk of other poor pregnancy outcomes, which often co-occur with preeclampsia, including preterm birth and fetal growth restriction.[71–73]

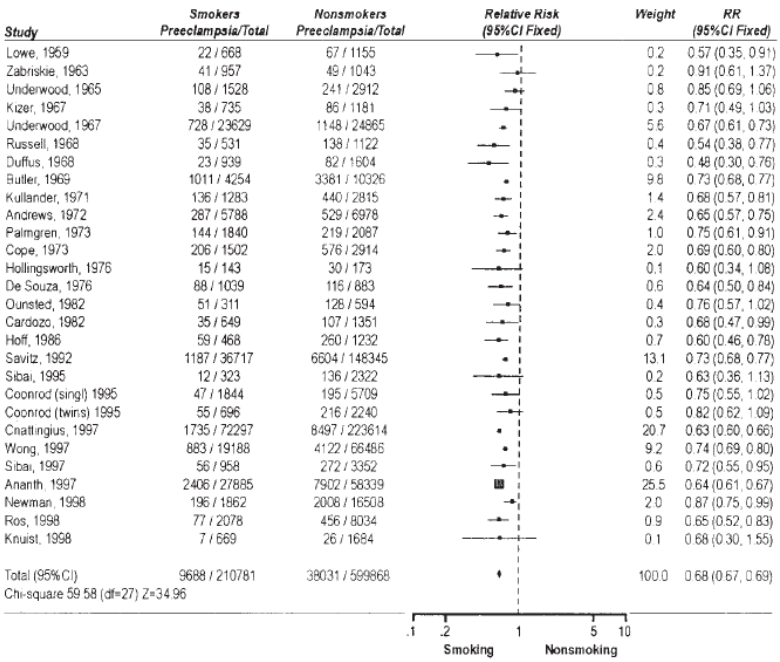


Figure 2. Association between cigarette smoking and preeclampsia in cohort studies. From Conde-Agudelo et al. 1999.[71]

Studies of smokeless tobacco (*snus,* used in Sweden), have not found a similar reduction in risk.[74] One can only make limited inferences the effect of timing of smoking on preeclampsia and dose-response relationships between cigarette smoking and preeclampsia because many women who smoke prior to pregnancy do not continue to smoke throughout the entirety of their pregnancy, and those that do tend to reduce their quantity of smoking. Some evidence indicates that women who smoke more cigarettes have the lowest risk of preeclampsia[71] and this risk may differ by timing of smoking during pregnancy.[75]

Biological Mechanism of Carbon Monoxide and Nitric Oxide

A biological mechanism explaining this inverse relationship between smoking and preeclampsia has been hypothesized but not definitively established, and appears paradoxical given the contradictory relationship between smoking and cardiovascular disease among non-pregnant adults and by the contradictory relationship of smoking with fetal growth restriction

and/or birthweight among normotensive pregnancies. One often posited explanation involves carbon monoxide and nitric oxide.[8,15,76] CO and NO have been shown to be associated with smooth muscle relaxation and blood pressure regulation[12,14], and there is evidence of their role specifically in the placental vascular system.[15,77] Because of these characteristics, CO and NO may be particularly relevant for preeclampsia.[15] Nitric oxide production and serum metabolites of NO are found to be lower among women with preeclampsia than during normal pregnancy,[16,22] and women with preeclampsia have been found to have decreased amounts of CO concentrations in their exhaled breath compared to those with healthy pregnancies.[59,78] The endothelium from chorionic and umbilical vessels releases NO,[79] and CO can be directly produced by trophoblasts.[80] It has been suggested that both NO and CO are required for trophoblast differentiation and invasion to occur properly.[8,15,77] Due to endothelial dysfunction, high vascular resistance and hypoxia may result from reduced endothelial-derived nitric oxide bioavailability,[17,22,81] contributing to hypertensive disorders of pregnancy.[82–86] Additionally CO has been shown to directly inhibit secretion of sFlt-1[59], resulting in an antiangiogenic state.

Pregnant women may be exposed to CO and NO endogenously or exogenously by inhaling cigarette smoke or air pollutants. Pathways influencing CO and NO production are highly integrated, and the synthesis of these gases occur primarily through the degradation of heme in hemoglobin and conversion of L-arginine (Figure 3). Catalyzed by heme oxygenase (HO-1, HO-2), heme breaks down into biliverdin, iron, and carbon monoxide. [11,87,88] Carbon monoxide then triggers nitric oxide synthase (NOS) of two types -- endothelial nitric oxide synthase (eNOS) and inducible nitric oxide synthase (iNOS).[14] Nitric oxide synthase then catalyzes the production of NO in the body from L-arginine, an amino acid synthesized in the body and found in dietary sources. NO, in turn, increases HO-1 expression.[89]

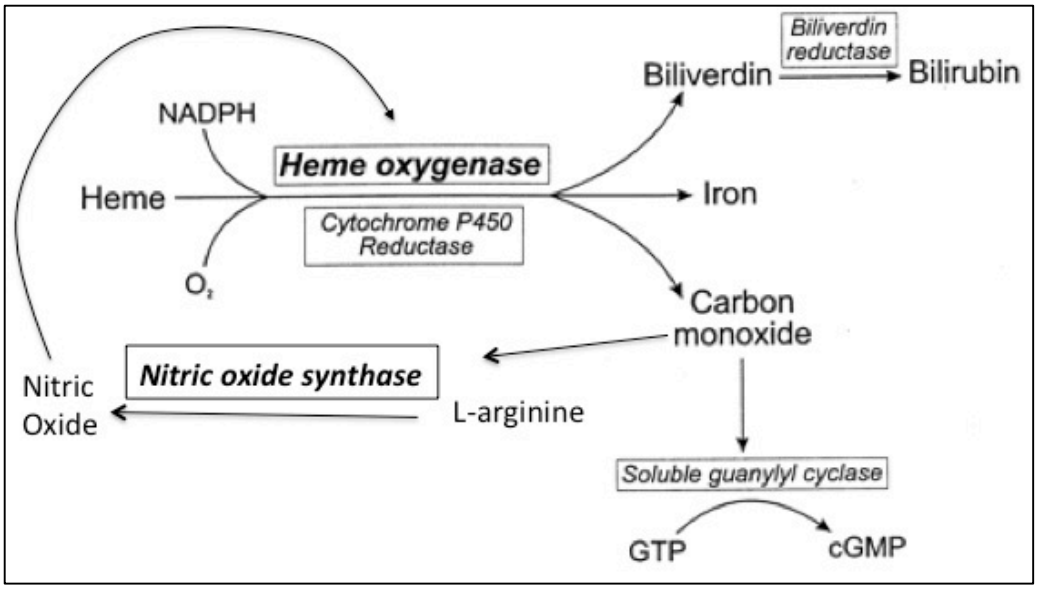


Figure 3. Nitric oxide and carbon monoxide signaling.

Animal models demonstrate that induction of HO-1 can reduce hypertension and the angiogenic imbalance characteristic of preeclampsia.[58,90] HO-1 protein expression is significantly lower in preeclampsia placentae than those with normal pregnancies.[14,80]
The role of carbon monoxide and nitric oxide may partially explain the enigmatic relationship between maternal smoking and preeclampsia. Smokers overall have reduced plasma levels of the anti-angiogenic factors sFlt-1, similar to the reduction found in women with preeclampsia, even after controlling for pregnancy outcome (779.6 pg/ml, 95% CI:487.5-1140.8 vs 1116.5 pg/ml, 95% CI: 793.6-1905.2).[91] Studies have shown that both exogenously administered CO and NO inhibit platelet aggregation and relax blood vessels by activating soluble guanylate cyclase and increasing levels of guanosine 3' 5' cyclic monophosphate (cGMP) as endogenously produced CO and NO would.[89] Smoking has the potential to modify the relationship between carbon monoxide signaling genes and preeclampsia, though this relationship could be synergistic or antagonistic. Most studies of the effect of cigarette smoke on

these pathways have assessed nitric oxide synthase activity in pulmonary endothelial cells and have found reduced eNOS and iNOS activity with cigarette smoke exposure.[92,93]

Biological Mechanism of Smoking Detoxification

The primary aims of the study are to investigate nitric oxide and carbon monoxide signaling pathways as a potential biological mechanism for the inverse association between smoking and preeclampsia, and to whether the effects of genes in smoking-related pathways are modified by exposure to smoking during pregnancy. Thus, in addition to CO and NO signaling pathways, we also include smoking detoxification genes and their relationship with preeclampsia.

Nicotine and polycyclic aromatic hydrocarbons are both components of cigarette smoke that activate receptors involved in smoking detoxification. Nicotine stimulates the nicotinic acetylcholine receptor and there is evidence in vitro that nicotine restores proangiogenic functions to endothelial cells harmed by soluble fms-like tyrosine kinase and soluble endoglin while stimulating placental growth factor,[94] all of which are associated with preeclampsia.[59,61,95] Polycyclic aromatic hydrocarbon response occurs through the aryl hydrocarbon receptor pathway (Figure 4), comprised of three phases (Figure 5). In phase I toxins are activated by enzymes such as those of the cytochrome P450 mono-oxygenase (CYP) family and epoxide hydrolase (EPHX). In phase II the reactive intermediate is conjugated to glutathione (GSH) by enzymes such as the glutathione S-transferases (GST). In phases III, these conjugates are less biologically active and more hydrophilic, so are then excreted via urine or bile.

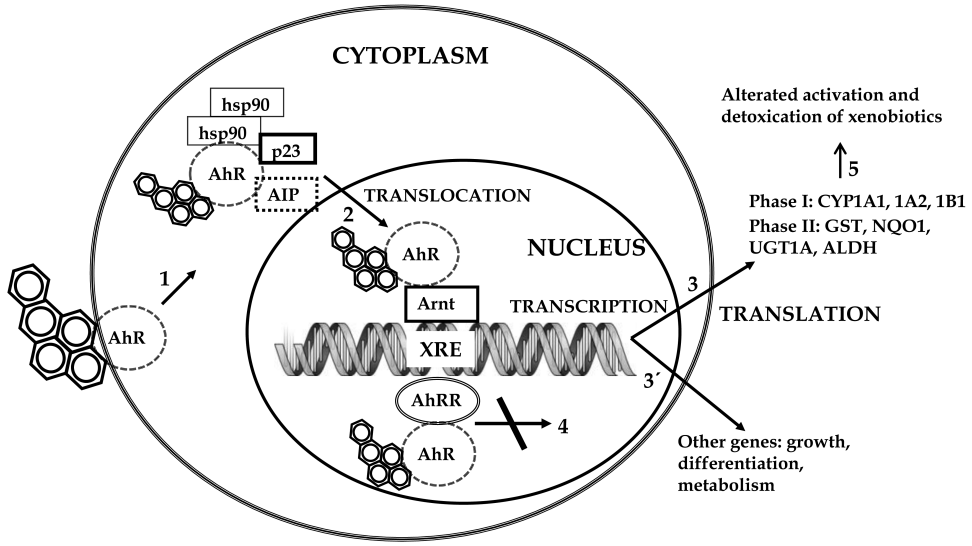


Figure 4. Aryl hydrocarbon receptor signaling and other related detoxification pathways[96]

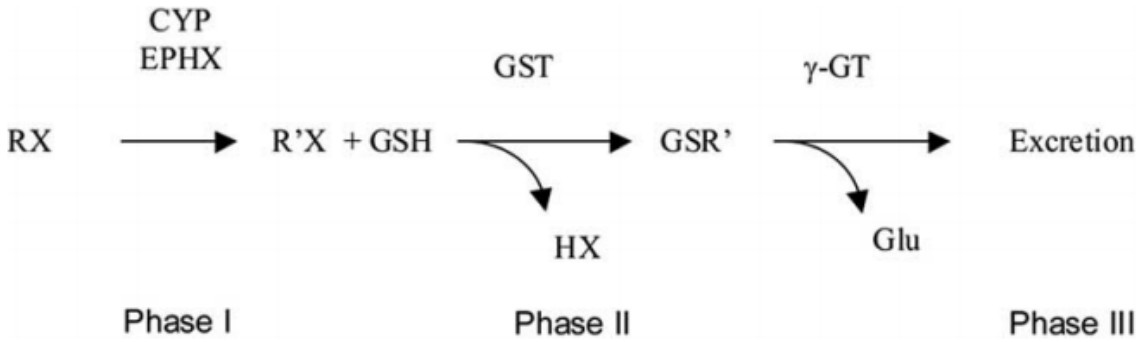


Figure 5. Overview of Phase I, Phase II, and Phase III xenobiotic metabolism[97]

Canonical Pathways

Below we describe the canonical pathways that served as the basis for selection of candidate genes involved in smoking detoxification and CO/NO signaling.

Three canonical pathways describe carbon monoxide and nitric oxide activity, and genes were grouped within these canonical pathways for analysis of aim 1. These pathways include endothelial nitric oxide synthase (eNOS) signaling, heme degradation, and hypoxia inducible

factor – 1 alpha (HIF1α) signaling. Five additional pathways were also included in aim 2 and these are xenobiotic metabolism, aryl hydrocarbon receptor signaling, glutathione-mediated detoxification, and nicotine degradation II, and nicotine degradation III.

eNOS Signaling

Nitric oxide is endogenously synthesized from L-arginine by nitric oxide synthases. Endothelial nitric oxide synthase is the isoform found in the endothelium and platelets, and is most relevant to preeclampsia for its role in vasodilation of the endothelium.[98] Other related isoforms are inducible nitric oxide synthase and neuronal nitric oxide synthase. Lower levels of eNOS expression have been found in syncytiotrophoblasts from the placentae of women with preeclampsia compared to those in placentae of women without preeclampsia despite no significant differences in levels of asymmetric dimethylarginine, which is an endogenous inhibitor of eNOS.[99]

Heme Degradation

Carbon monoxide is endogenously synthesized in heme catabolism. Heme, originating primarily from hemoglobin, is cleaved by heme oxygenases to form biliverdin, then bilirubin and carbon monoxide. Heme oxygenase occurs as 2 isozymes, inducible heme oxygenase (HMOX1) and constitutive heme oxygenase (HMOX2).[100] In studies of placental vascular formation in mice as determined by micro-computerized tomography, mice with a partial deficiency in HMOX1 were found to have malformations of the fetomaternal interface and insufficient spiral artery remodeling.[101]

HIF1α Signaling

Hypoxia inducible factor 1 is a transcription factor that regulates oxygen homeostasis.[102–104] It regulates response to hypoxia by activating transcription of genes whose protein products

increase oxygen delivery or facilitate adaptation to hypoxic conditions,[102] and is important in embryonic and placental vascularization and pathophysiology of ischemic disease, such as preeclampsia.[102] Because of the role in oxygen adaptation and delivery, HIF1A and related genes are heavily involved in both carbon monoxide and nitric oxide signaling. HIF1A is overexpressed in placentas from women with preeclampsia,[105] and cell free plasma expression of HIF1A is greater in pregnancies complicated by hypoxia and/or intrauterine growth restriction than in uncomplicated pregnancies.[106] Additionally, in a placental expression study of 57 women terminating pregnancies at 6-8 weeks' gestation, expression of HIF1A in active smokers was found to be significantly higher than non-smokers (estimate not provided, p=0.003)[107] indicating the potential importance of the effect of smoking on the placenta during early pregnancy, particularly as it relates to preeclampsia.

Xenobiotic Metabolism

Xenobiotics, such as those that are contained in cigarettes trigger cellular stress response, leading to cell proliferation, differentiation, apoptosis or necrosis.[108] The xenobiotic metabolism pathway induces expression of enzymes to protect the body from these harmful responses. These enzymes include phase I enzymes (CYP, ALDH, FMO) that usually create a polar group, phase II enzymes (UGT, GST, SULT) that produce hydrophilic products, and phase III enzymes (MDR1, OATP2, MRP) which are transporters that export xenobiotics and their products out of the cell.[108] The xenobiotic metabolism pathway is highly integrated with the aryl hydrocarbon receptor signaling and nicotine degradation pathways. Gene by environment interactions have been studied among some of the genes in these pathways and these associations are summarized in the section of SNPs in genes involved in smoking detoxification.

Aryl Hydrocarbon Receptor Signaling

Halogenated and polycyclic aromatic hydrocarbons are contained in cigarette smoke.[108] Figure 5 describes the aryl hydrocarbon receptor signaling pathway, in which the aryl hydrocarbon receptor (AhR) mediates xenobiotic metabolism, dioxin and polycyclic aromatic hydrocarbon toxicity, and vascular development.[109] AhR is highly expressed in the placenta[110] and higher in preeclamptic pregnancy placentas than normal pregnancy placentas.[111] AhR has been shown to increase invasion of cancer cells in several types of cancer, and it is hypothesized that AhR may increase invasion of fetal cytotrophoblast cells as well, protecting against preeclampsia.[111]

Glutathione-mediated Detoxification

Figure 5 describes the general detoxification pathway, where the first step can be catalyzed by many different enzymes, collectively known as glutathione transferases, or GSTs. In the first reaction, R may be an aliphatic, aromatic or heterocyclic group, and X may be a sulfate, nitrile or halide group. General Background Thiols play several major roles in the cell; they help maintain the redox balance, keeping a reduced environment (see the pathway glutathione redox reactions II), they fight reactive oxygen and nitrogen species (ROS and NOS, respectively), and they are involved in the detoxification of many other toxins and stress-inducing factors. In most organisms the major thiol is the tripeptide glutathione (γ-Glu-Cys-Gly, known as GSH). GSH is active against toxins by a process that involves multiple enzymes, and in the case of eukaryotes, occurs across multiple organs.

Nicotine Degradation II and III

Nicotine is the principal alkaloid in the leaves of tobacco and is available in both smokeless and smoking tobacco. In cigarette smoke, it is carried into the body on particulate matter. It binds to the nicotinic cholinergic receptors and activates dopaminergic reward system

in the brain. It is primarily metabolized in the liver and most is metabolized to cotinine. It is metabolized primarily through signaling of the CYP enzymes as with other xenobiotics.[112]

Analogy with Air Pollution

Some have suggested that the mechanism for smoking and preeclampsia should mirror associations seen with other tobacco or combustible exposures, but studies of air pollution or environmental tobacco smoke and preeclampsia have not shown consistent results and generally do not find the similar reduction in risk found with smoking studies. Most studies of air pollution have measured particulate matter, nitrogen oxides, carbon monoxide and ozone at air monitors of closest proximity to study participants.[113–121] Four studies specifically analyzed either CO or NO; of these, three found an increased risk of preeclampsia for both CO and NO[117–119] and one found a decreased risk of preeclampsia with carbon monoxide exposure.[115]

Though one may expect a similar inverse relationship between air pollution and preeclampsia because of the gaseous composition, differential results of smoking studies and air pollution studies may exist for several reasons. Exposure from the ambient environment may have a different effect than personal exposure to cigarettes. In particular, there has been no relationship found between second-hand smoke from partner smoking and preeclampsia or pregnancy induced hypertension[122] which may indicate differences in delivery or dose of the exposure. Additionally, the studies of air pollution described above have all used registry data for risk estimates, which have a greater potential of outcome misclassification and likely encompass a large number of women with pregnancy induced hypertension rather than preeclampsia.

C. Genetics of Preeclampsia

Familial Associations

Numerous studies have suggested a familial predisposition for preeclampsia risk in first-degree relatives of women with preeclampsia [123]. In studies of familial aggregation of preeclampsia, genetic factors may contribute more than 50% of heritability [124–126], which is supported by both large-scale registry-based epidemiologic studies[127,128] and studies of transgenic mice.[123,129] A number of candidate regions have been identified through linkage studies[124], and approximately a dozen genes show altered placental expression in preeclampsia in at least two independent studies.[123]

Early Linkage and Candidate Gene Studies

A search was done on PubMed as well as the Preeclampsia SNP Database (PESNPdb)[130] and The Genetic Association Database[131] for previously found associations of SNPs with preeclampsia (Table 1). Although there have been hundreds of genetic studies of preeclampsia, including placental expression, linkage analyses, candidate SNP, and genome wide association studies, findings have been inconsistent.

Some of the most replicated relationships in the genetic preeclampsia literature are among linkage analyses detecting potential susceptibility loci. Genome wide scans have detected associations with preeclampsia of variants on chromosome 2 (on 2p11-14 and 2q22-23) in Icelandic (LOD score=4.70)[132], Australian and New Zealand (LOD score=2.58)[133], and Finnish populations (NPL score=3.77, p=0.000761)[134]. Roten et al. did a scan of 71 SNPs in these regions of chromosome 2 in Norwegian women and found four significant SNPs in *ACVR2A* using a false discovery rate of 5%.[135] A genome wide association study found two variants meeting modified Bonferroni significance (rs7579169, $p=3.58 \times 10^{-7}$ and rs12711941, $p=4.26 \times 10^{-7}$) in an intergenic region near *INHBB* on 2q14.2.[136] Laivuori et al. found an

additional susceptibility locus in 9p13 in the same Finnish population, but this has not been replicated in other studies.[134]

Among the more recent and larger candidate gene studies (those with at least 150 cases (Table 1)), few variants have been consistently associated with preeclampsia. This is partially due the selection of different markers; for some there has been no attempt at replication apart from through GWA studies. However, among the better powered candidate gene studies, some replication has been found for variants within *TNFA*[137–139] and *IL10*,[124,138,139] although there are also studies that have found no associations.[140,141]

Buurma et al. conducted a recent meta-analysis of genetic variants and preeclampsia, finding 542 genetic associations of preeclampsia with 22 replicated genetic variants.[142] Of these 22, only seven variants were significantly associated with preeclampsia following meta-analysis. These are: *FV* rs6025, *FV* rs6020, *F2* rs179963, *SERPINE1* rs1799889, *ACE* rs4646994, *CTLA4* rs231775, and *LPL* rs268.[142] The majority of the studies considered, however, are quite small (range=7 to 665 cases, median=102, mean=119) which is a possible reason for unreplicated findings and lack of associations after meta-analysis. For example, for variants in *TNFA* and *IL10*, which were those associations most replicated in studies with at least 150 preeclampsia cases, the number of cases were small: 1075 cases among 8 studies for *IL10* and 1592 and 390 cases for two variants of *TNFA*, respectively.[142]

In addition to the lack of overlap in typed markers and small study sample sizes, several other factors may influence the inconsistent findings in candidate gene studies of preeclampsia. It is plausible there is truly no effect of any variants on preeclampsia aside from those described in the meta-analysis. However, it is also plausible that substantial heterogeneity among these studies make comparisons across studies difficult. Differences in study design are described below.

Although all of the studies used preeclampsia as an outcome, there were a variety of definitions used. For example, some of the studies only included women with severe

preeclampsia.[143–147] In older studies, a change in blood pressure rather than a set cutpoint was often used as a criterion for preeclampsia.[148–151] Because risk factors differ for various subtypes of preeclampsia,[62] and different presentations of preeclampsia may have a different underlying etiology, the inclusion or exclusion of various subtypes could yield conflicting results.

The method of case ascertainment may affect validity and replicability. Nearly all of the candidate gene studies are case-control studies, and study participants were generally recruited or selected from hospitals or obstetric clinics. Preeclampsia diagnosis was almost always determined from clinical records and hospital chart review, which can reduce the likelihood of misclassification, though misclassification may still occur due to lack of multiple blood pressure and urinary measurements over time and differential practices of health care providers. Misclassification of the outcome would not likely bias results if non-differential,[152] however, could increase noise and reduce power to detect an effect.

Lastly, the study populations of all the preeclampsia genetic studies are extremely diverse, with populations from countries including India[153], Korea[99], United States[154], Brazil[83], and Colombia[155], among others. Some of the individual studies are done in places like the United States, in which the population is ethnically heterogeneous, which opens up the potential for population stratification bias within the study.[156] When making comparisons across studies, populations of different ancestry may show different associations. There is variation in LD among populations of different origins,[157] which can lead to different SNP selection, haplotypes, and associations for different populations.

Genome Wide Association Studies

PubMed as well as the National Human Genome Research Institute Catalog of Published Genome Wide Association Studies was searched for GWA studies of preeclampsia. To date, there are only three GWAS of preeclampsia, and none have found any associations meeting Bonferroni significance.[136,158,159] The first genotyped 648,175 SNPs in 538 women with

preeclampsia and 540 with normal pregnancies of Caucasian ancestry in Australia. Two SNPs in an intergenic region near *INHBB* on chromosome 2 met their modified significance threshold (rs7579169, $p=3.58 \times 10^{-7}$ and rs12711941, $p=4.26 \times 10^{-7}$).[136] Another genome-wide scan of 705,969 SNPs in 177 preeclampsia cases and 116 normotensive controls was conducted among mothers who gave birth in Iowa from Aug 2002 to May 2005 as part of the SOPHIA study. The top four SNP associations had genotypic p-values between 10^{-5} and 10^{-6}.[158] The third preeclampsia GWAS is another study by Zhao and colleagues, a subset of the Hyperglycemia and Adverse Pregnancy Outcome study cohort of a diverse group of mothers from Barbados; Brisbane and Newcastle, Australia; Toronto, Canada; Belfast, United Kingdom; Bellflower, California; and Bangkok, Thailand. This study analyzed approximately 979,693 SNPs and the study population consisted of Afro-Caribbean women (21 cases, 1010 controls), Hispanic women (62 cases, 658 controls), and European ancestry women (50 cases, 1202 controls).[159] As in the other cohorts, none of the variants were Bonferroni significant, though the top SNP for each ethnic group had p-values between 10^{-6} and 10^{-7}.[159] None of the top SNPs were replicated in other studies. The closest replication of variants were those from the HAPO study that were "suggestively replicated" in the SOPHIA study; two variants in an intronic region of *INVS* among women of Caucasian ancestry.[159]

Although the GWA studies are generally larger than the candidate gene studies, they all still face the challenge of being underpowered when considering multiple comparisons for the number of SNPs analyzed. In particular, the Iowan study of 705,969 SNPs in 177 preeclampsia cases and 116 controls[158] is far too small to detect any reasonable effects. This study by Zhao and colleagues also leaves the greatest potential for outcome misclassification of the three GWA studies. Potential cases were first identified through electronic birth certificates that were "check box positive" for either pregnancy induced hypertension or eclampsia.[158] It is quite possible that a woman could not have had the box checked and still met the clinical criteria for preeclampsia. In an effort to ensure the validity of cases and controls, cases were confirmed by

telephone interviews with study participants and the medical record, however, evidence has shown that validity of self-reported preeclampsia is moderately poor. In a validation study as a part of the Generation R cohort study in The Netherlands, 50% of women who self-reported preeclampsia (72/152) were determined not to meet the clinical criteria for disease.[160]

Of these studies, the one by Johnson et al. seems the strongest. It has the largest number of cases by approximately three-fold (n cases = 538), a fairly homogeneous population of Australian women of Caucasian ancestry (PCA revealed minimal population structure), and case ascertainment determined from new-onset hypertension and protein levels confirmed by medical records.[136] This study, however, is still likely to be underpowered at all but the most common allele frequencies.

These three GWA studies have not found any significant associations at genome-wide significance levels, however, even the top hits that may be suggestive of association are different in each study. This may be, in part, due to the problem that these three studies are not necessarily good comparisons for one another. Although each of them have the greatest number of women of European ancestry compared to other ancestral groups within each study and all were assessed for population stratification using principal components analysis, the HAPO population is quite heterogeneous. Although it has the largest sample size (though not the largest number of cases), the sample is divided among three different ancestry groups (Afro-Caribbean: 21 cases, 1010 controls; European ancestry: 50 cases, 1202 controls; Hispanic: 62 cases, 658 controls).[159] These three studies also do not use the same case definition for preeclampsia. Although they all consider preeclampsia to be de novo hypertension of a systolic blood pressure of at least 140 mmHg and/or diastolic blood pressure of at least 90 mmHg in the presence of urinary protein, the study by Johnson et al. includes a change in systolic blood pressure of 25mmHg and/or diastolic blood pressure of 15 mmHg as diagnostic criteria.[136] Each of the studies also uses different criteria for proteinuria (Johnson et al. 2012: ≥0.3 g/l in a 24 hr specimen or 2+ dipstick reading or spot protein:creatinine ratio ≥0.3 g/mmol[136]; Zhao et al. 2012:

1+ dipstick reading from two or more specimens collected at least 4 hours apart, one or more dipstick values of 2+ near end of pregnancy, one or more catheterized dipstick values of 1+ during delivery hospitalization, or 24 hr urine collection with protein >300mg[158]; Zhao et al. 2013: dipstick protein value ≥1+ or 24 hr urine collection with protein ≥300 mg[159]). These differences in population and case definition decrease the ability to replicate findings, particularly in underpowered studies.

Because of the lack of replicable findings among GWAS of preeclampsia, some researchers are beginning to use methods that may account for a potential polygenic effect rather than a single SNP effect. Smith et al. use genome wide data to investigate genetic risk scores for essential hypertension and their association with preeclampsia.[161] They found that genetic risk scores for hypertension were not associated with a risk of preeclampsia, suggesting that a different underlying genetic etiology of preeclampsia and essential hypertension.

Table 1. Genetic studies of preeclampsia with 150 cases or more

Study	Population	Study Design	Cases	Controls	Sample	Genes	Association
O'Shaughnessy, KM et al. 1999[161]	Women recruited at preeclampsia diagnosis from East Anglian region of the United Kingdom	Case-control	283 cases; 149 had severe PE, 25 developed HELLP	100 age-matched from same hospital	Maternal blood	*MTHFR* (C677T variant, MAF=0.03) *F5* (G1619A variant, MAF = 0.32)	None None
De Groot, CJ et al.1999[162]	Primigravid women recruited by computer database and patient charts from 2 hospitals in The Netherlands	Case-control	163 cases	163 controls	Maternal blood	*F2* *F5*	None None
Rajkovic, A et al. 2000[163]	Black African women at Harare Maternity Hospital in Zimbabwe	Case-control	171 cases	185 normotensive controls	Maternal blood	*MTHFR* (C677T variant, MAF=0.09)	None
Kim, YJ et al. 2001[164]	White patients from University of Iowa Hospitals and Wake Forest University between 1996 and 1999	Case-control	281 cases	360 controls	Maternal blood or buccal cells	*MTHFR* (C677T variant, MAF=0.32) *F5* (G1691A variant, 0.02)	None None
Kim, YJ et al. 2001[165]	White patients from University of Iowa Hospitals and Wake Forest University	Case-control	250 mothers, 106 offspring	265 controls	Maternal and child blood or buccal cells	*LPL* (3 polymorphisms)	None

Study	Population	Study Design	Cases	Controls	Sample	Genes	Association
Lachmeijer, AMA et al. 2001[136]	Families of affected sisters and their parents from discharge records, obstetrical charts and advertisements in The Netherlands between June 1995 and Oct 1997	Family based	150 sibling pairs	98 men and women from Vrije University	Blood	*TNFa* and *LTA* (5 haplotypes)	TNF-1 haplotype association, OR 1.9 (95% CI: 1.1, 3.3)
Lachmeijer, AMA et al. 2002[166]	Families of affected sisters and their parents from discharge records, obstetrical charts and advertisements in The Netherlands between June 1995 and Oct 1997	Family based	150 sibling pairs	104 healthy blood donors (male and female)	Blood	*IL1B* (2 SNPs) *IL1RN* (IVS2 variant)	None None
Levesque, S et al. 2003[139]	French Canadian nulliparas being investigated for development of hypertensive disorder in Quebec City between 1996 and 2000	Case-control	180 cases	310 matched controls	Maternal blood	*GSTP1* (Ile105Val, MAF=0.11) *TGFB1* (Arg25Pro, MAF=0.19) *TNF* (G-308A, MAF=0.13) *AGT* (Thr174Met, MAF=0.08) *AGTR1* (A1166C, MAF=0.33) *F5* (Gin506Arg, MAF=0.03) *APOB* (Thr2488Thr, MAF=0.5)	None None None p = 0.003 None None None
Malina, AN et al. 2004[167]	Women delivering at Magee Women's Hospital in Pittsburgh, PA	Case-control	177 cases	179 nomotensive controls	Maternal blood	*ADRB3* (Trp64Arg, MAF=0.09)	None

Study	Population	Study Design	Cases	Controls	Sample	Genes	Association
Haggerty, CL et al. 2005[137]	Primiparous women from the PEPP study at Magee Women's Hospital in Pittsburg, PA	Case-control stratified by race	150 cases (20 Black, 130 White)	661 nomotensive controls (199 Black, 462 White)	Maternal blood	*TNFa* (1 SNP)	OR=4.1 (95% CI: 1.1, 15.3) among white women
						IL1A (2 SNPs)	OR=11.6 (95% CI: 1.5, 89.3) among black women
						IL1B (1 SNP)	Haplotype with IL1A association in all women
						IL10 (2 SNPs)	OR=1.9 (95% CI: 1.0, 3.9) among white women
Goddard, KA et al. 2007[123]	Mothers and infants from a hospital based population in Puerto Alto, Chile	Case-control	394 mothers, 324 offspring	602 mothers, 631 offspring	Maternal blood and child cord blood	190 genes, 775 SNPs	
						IL1A	p=0.0014
						IL12RB1	p=0.0093
						IL4R	p=0.0006
						IL1B	p=0.0272
						IL10	p=0.045
						TGFb1 *NOS3* *MTHFR*	None None None
						IL6	None
						IFNgamma	None
						HL	None
						CAPN-10	None

Study	Population	Study Design	Cases	Controls	Sample	Genes	Association
Mirahmadian, M et al. 2008[138]	Consecutive patients at 2 hospitals in Tehran, Iran between July 2006 and July 2007	Case-control	160 cases	100 healthy pregnancies during same time period	Maternal blood	*TNFa* (2 SNPs)	p=0.0001
						IL10 (3 SNPs)	p=0.001
						TLR2	None
						MMP9	None
						IL6	None
						IL6	None
Johnson, MP et al. 2009[140]	Australian Family Cohort and Norwegian Case Control (from HUNT cohort and biobank)	Family based and case-control	74 Aust/NZ families with 140 affected women, Norway 1,139 women	Aust/NZ 90 unaffected, Norway 2,269 controls	Maternal blood	10 genes (56 SNPs) within a previously identified 5q critical region	
						IL4 (3 SNPs)	None
						CSF2 (2 SNPs)	None
						ERAP1 (2 SNPs)	p=0.009
						ERAP2 (2 SNPs)	p=0.004
						IL13 (7 SNPs)	None
						IL3 (1 SNP)	None
						IL5 (2 SNPs	None
						IL1B	None
						IL10 (3 SNPs)	None
						TNFa (3 SNPs)	None
						LGALS13 (10 SNPs)	None

Study	Population	Study Design	Cases	Controls	Sample	Genes	Association
Fenstad, MH et al. 2010[168]	Families from Australia and NZ, mothers from Norway	Family linkage design and case-control	74 Aus/NZ families with 14 affected women and 851 women in Norway	Aus/NZ 146 unaffected, 1440 controls in Norway	Maternal blood	*TNFSF13B* (7 SNPs Aus/NZ, 3 SNPs Norway)	3 rare SNPs p=0.0153 in Aus/NZ families; Not replicated in Norwegian
Hill, LD et al. 2011[169]	Mothers and infants from Santiago Chile, Philadelphia PA and Detroit MI	Case-control	Chilean: 528 dyads, African American (unpaired: 424 mothers, 375 infants	Chilean: 575 dyads, African American (unpaired): 412 mothers, 462 infants. All term	Maternal blood, cord blood or neonatal cheek swabs	*ERAP2* (rs2549782, rs17408150)	p=0.009 for fetal rs2549782 in African Americans
Johnson MP et al. 2012[135]	Women restrospectively ascertained from larger case-control cohort recruited at Royal Women's Hospital, Melbourne Australia 2007 to 2011. Women of Caucasian ancestry	Case-control	538 preeclampsia cases	540 normal pregnancy controls	Maternal blood	648,175 SNPs rs7579169, MAF=0.34, intergenic near *INHBB* rs12711941, MAF=0.34, intergenic near *INHBB*	None passed Bonferroni correction. $p=3.58 \times 10^{-7}$ $p=4.26 \times 10^{-7}$
Zhao L et al. 2012[157]	3078 primiparous mothers who gave birth in Iowa from Aug 2002 to May 2005; Recruited through SOPHIA study; Identified through electronic birth certificates check box positive for PIH or eclampsia	Case-control	177 PE Cases confirmed by telephone interview and medical record	116 normotensive controls	Maternal blood	705,969 SNPs rs1426409, MAF=0.31, intergenic near *KIAA1239* rs1426409, MAF=0.31, intronic *ESRRG* rs9831647, MAF=0.48, intronic *LMCD1* rs10743565, MAF=0.43, intronic *IFLTD1* Copy number deletion in PSG11	None passed Bonf correction of 7.1 x 10-8 $p=3.14 \times 10^{-6}$ $p=3.80 \times 10^{-6}$ $p=9.36 \times 10^{-6}$ $p=1.64 \times 10^{-5}$

Study	Population	Study Design	Cases	Controls	Sample	Genes	Association
Zhao L et al. 2013[158]	Afro-Caribbean, Hispanic, and European ancestry women from the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study	Case-control	Afro-Caribbean 21 cases; Hispanic 62 cases; European 50 cases	Afro-Caribbean 1010 controls; Hispanic 658 controls; European 1202 controls	Maternal blood	979,693 SNPs for Afro-Caribbean, 964,533 for Hispanic; 541,023 for European met criteria (MAF>0.01) rs11617740 in *FGF14* rs17412740, intergenic near *LZTS1* rs7322722 in *MYCBP2* *FGF14* Copy number variants	None significant after Bonferroni correction $p=7.32 \times 10^{-7}$ in Afro-Caribbean $p=2.14 \times 10^{-6}$ in Hispanic $p=2.93 \times 10^{-6}$ in European ancesty $p=7.32 \times 10^{-7}$ in Afro-Carribean $p=2.14 \times 10^{-6}$
Wan JP et al. 2014[170]	1218 Northern Han Chinese Women; excluded for chronic HTN, multiple pregnancies, previous renal autoimmune, metabolic CVD, loss to follow-up	Case-control	515 PE cases, second stage 166 early-onset	703 healthy controls, second stage 178 controls	Maternal blood	rs2681472, MAF=0.36	p=0.009 for early-onset PE
Wan JP et al. 2014[171]	Han Chinese pregnant women who visited Provincial Hospital, Dec 2009 to Nov 2011.	Case-control	454 PE cases (149 mild, 305 severe; 211 early-onset, 243 late-onset)	460 normotensive controls	Maternal blood	rs11646213, MAF=0.14	p=0.017 for PE overall p=0.002 for severe PE p=0.004 for early-onset PE

Study	Population	Study Design	Cases	Controls	Sample	Genes	Association
Smith et al. 2016[160]	Iowa residents part of the Study of Pregnancy Hypertension in Iowa (SOPHIA). Nulliparous residents who had a live birth from Aug 2002 to April 2005. Replication from 5 US sites. European ancestry.	Case-control	162 PE cases (516 PE cases in the replication sample)	108 normotensive controls (1097 controls in the replication sample)	Maternal blood, saliva, buccal samples	Hypertension GRS Systolic blood pressure GRS Diastolic blood pressure GRS Mean arterial pressure GRS	None associated with PE
Thomsen et al. 2017[172]	Women in the Norwegian HUNT study. 1984 to 2008 in Nord-Trondelag. Primarily northern European origin. Replication cohort from Australia Royal Women's Hospital, Melbourne.	Case-control	1006 PE cases (471 PE cases in the replication sample)	816 nonpreeclamptic controls (547 controls in the replication sample)	Maternal blood	rs17367504, MAF=0.13	0.65 (0.53 – 0.80), p=3.52 x 10^{-5}

Single Nucleotide Polymorphisms in Genes Involved in CO and NO Activity

Genetic factors play a role in how carbon monoxide and nitric oxide are processed and synthesized in the body, and these genes have reported associations with hypertensive disorders. Each of the enzymes involved in production of NO and CO described earlier are the product of a specific gene: *HMOX1* codes for heme oxygenase 1, *NOS1* codes for neuronal nitric oxide synthase, *NOS2* codes for inducible nitric oxide synthase, and *NOS3* codes for endothelial nitric oxide synthase. Nitric oxide is not stored after synthesis and has a short half-life in tissue, so production is partially regulated through alterations in expression or activity of eNOS.[162] The Glu298Asp[163], intron 4 [17], and -786T>C [17] polymorphisms of *NOS3* are associated with reduced basal NO production and low plasma NO levels and are also the most widely studied in relation to preeclampsia, as they are considered clinically relevant.[164] In mice, expression of HO-1 can be induced by the signaling molecule adenosine 3',5'-cyclic monophosphate (cAMP), which then stimulates formation of CO and NO.[165]

Several genes involved in our pathways have demonstrated associations with cardiovascular disease and hypertension. *NOS3* knockout mice are hypertensive.[166] A meta-analysis of human studies found *NOS3* polymorphisms to be associated with coronary heart disease (OR 1.17; 95% CI 1.07-1.28).[17] Similarly, SNPs of *NOS1* were associated with stroke susceptibility.[167] However, GWA studies have not found cardiovascular disease associations with any of the NOS genes.

Variants in *HMOX1* have also been associated with hypertension in several populations[19,168], though no associations of *HMOX1* have also been found in GWA studies of hypertension. Although NOS and HMOX genes are the most commonly studied, variants of seven other genes in our pathways have reported associations with cardiovascular disease phenotypes in GWA studies. These include *CAV1*[169,170], *ESR1*[171], *GUCY1A3*[172–174], *GUCY1B3*[172–174], and *PRKCA*[175] in the nitric oxide synthase pathway and *EDN1*[176] and *EPO*[177] in the HIF1A signaling pathway.

In preeclampsia specifically, as with other variants, those within *NOS3* have been widely studied but associations remain inconsistent (Figure 6), which as with genetic studies of preeclampsia overall, may be partly attributable to small sample sizes, heterogeneity in race/ethnicity of population, and differences in case definition. In a linkage study of 42 women with pregnancy induced hypertension in 50 families in Scotland and Iceland, Arngrimsson et al. found evidence of linkage of region 7q36 (LOD score=3.36), where *NOS3* is located, and a transmission disequilibrium test provided evidence of association for a marker in intron 13 of *NOS3* within this region (-2LL difference = 6.43, p=0.005).[178] Linkage to this region in other genome-wide scans, however, has not been confirmed in other populations.[133,134]

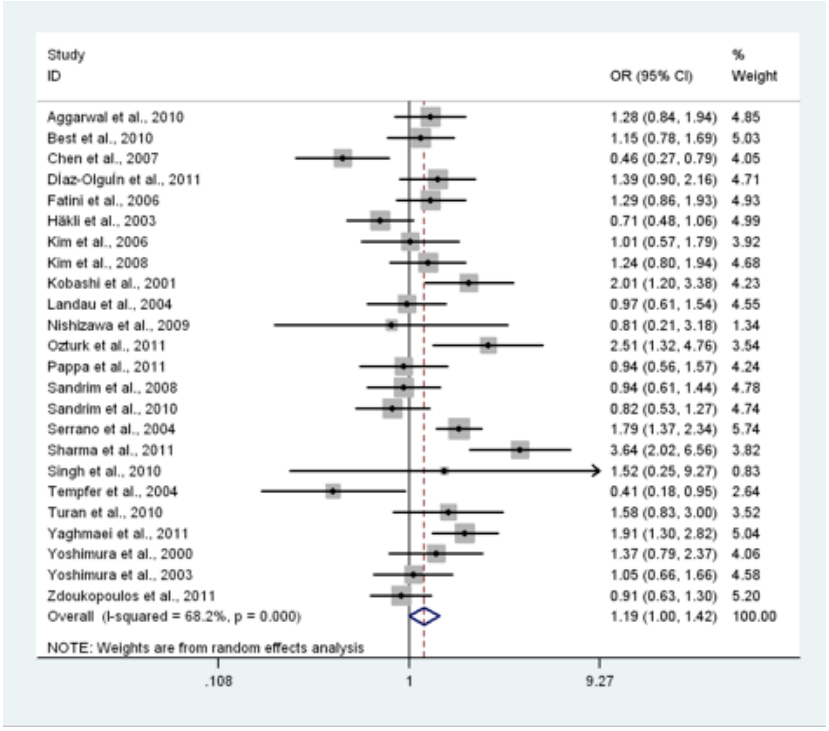


Figure 6. Forest plot of *NOS3* rs1799983 associations from Buurma et al. meta-analysis, supplementary data.[142] Figure references.[23,83,85,99,149,153–155,179–185]

There have been numerous candidate gene studies of *NOS3* and preeclampsia, reporting widely variable estimates on both sides of the null. A recent meta-analysis of the three most widely studied polymorphisms in *NOS3* (G894T, T786C, and intron 4b/a) found that only the G894T polymorphism in a recessive model was associated with preeclampsia after meta-analysis.[164] The recent meta-analysis by Buurma et al. analyzed findings of the variants 27 bp-VNTR in intron 4, rs2070744, and rs1799983 of *NOS3* (Figure 6). The study included only genetic variants that were associated with preeclampsia and for which independent replication was available. After meta-analysis of 29 studies of those variants, however, the authors found no association between any of these variants and preeclampsia.[142]

Single Nucleotide Polymorphisms in Genes Involved in Smoking Detoxification

Genes encoding interleukins (*IL1A,*[138,186] *IL1B,*[187] *IL4*[136]) and tumor necrosis factor (*TNF*)[137,138,186,188] are the most widely studied genes within the smoking detoxification pathways for their association with preeclampsia. Studies have shown generally null and unreplicated findings. The cytochrome P450 genes, *CYP1A1* and *CYP19A1,* have been studied primarily in Japanese populations; findings have been null.[189,190]

Smaller candidate gene studies of *CYP1A1*, *GSTT1,* and *GSTP1* have revealed potential interactions with exogenous exposures, for other reproductive outcomes including *CYP1A1* and benzene with shortened gestation, *GSTT1* and smoking with birthweight, and *GSTP1* and smoking with spontaneous abortion.[191]

Summary of Limitations of Existing Genetic Studies

There have been a large number of genetic studies of preeclampsia, however, these studies are limited in what we can determine about the etiology of the disease. The majority of genetic studies of preeclampsia are early linkage analysis studies, candidate gene studies, or placental expression studies.

Linkage analyses use information from related individuals and can provide a starting ground in which to conduct further analyses or select potential candidate genes by identifying potential regions of the genome that may be associated with disease[192]. With regards to the preeclampsia literature, potential susceptibility loci determined by linkage analysis have been some of the most widely replicated findings. Preeclampsia is a heterogeneous phenotype, in which unclear case definition can lead to misclassification of affected and unaffected individuals. In a linkage study, some selection criteria are usually applied to the phenotype before families are selected.[192] In preeclampsia, this makes the case definition consistent across related individuals, which may increase the ability to identify a potential linkage region. This, however, may reduce the ability to generalize the findings to the larger population.[192] Because linkage studies are also studies of related individuals, they likely share more environmental factors. Although it seems plausible that are shared environment may improve the ability to replicate genetic effects by excluding environmental effects, a simulation study by Tiret and colleagues found that ignoring a GxE interaction effect among nuclear families decreases power to detect a gene effect and biases parameter estimates.[193] Overall, linkage studies of preeclampsia can be a good starting place, however, they do not give the ability to find specific SNPs or determine anything about variant function, nor account for any environmental factors that may be important particularly for preeclampsia.

The majority of preeclampsia genetic studies are candidate gene studies, and there have been widely variable associations found for many of the genes studied. Although genes are generally selected based on biologically plausible hypotheses, past experience in genetic

epidemiology has demonstrated that scientists generally do a poor job of selecting candidate genes.

For gene expression studies of preeclampsia, placental tissue is typically used. In these studies, the placenta is collected after birth, therefore, it is collected and analyzed after the onset of disease. Although these studies can detect differences in placentas of women who had preeclampsia and those who have not, a temporal relationship cannot be determined. It is possible that expression differences may result from the consequences of preeclampsia rather than explain the onset of preeclampsia.

Of the three genome wide association studies of preeclampsia, none have found any genome-wide significant associations when correcting for multiple tests using the standard Bonferroni correction. In general, GWAS of preeclampsia are hugely underpowered to detect modest effects. Although the GWA studies are some of the largest preeclampsia studies, sample sizes have hovered in the hundreds of cases and controls rather than the thousands or tens of thousands necessary to be truly powered to detect an effect. Additionally, because preeclampsia is a heterogeneous phenotype, in which all subtypes of preeclampsia may not be associated with a single variant, small numbers within subtypes may make it even more challenging to detect an effect. With other phenotypes such as cardiovascular disease or cancer, in which the GWAS literature is more advanced, consortia are created to pool large cohorts of data to improve power. A consortium for preeclampsia has recently been established, InterPregGen, although no findings from this study have yet been published.

Small sample size and low power are both a problem in the smaller candidate gene studies and larger GWA studies of preeclampsia. To my knowledge, even among the 500+ studies of preeclampsia, only those in Table 1 have at least 150 cases of preeclampsia in the study population. The populations studied are highly diverse, representing ethnicity of over 20 countries. Reduction in power due to additional random noise in the population is also an important consideration that may reduce the ability to detect a genetic effect. Misclassification of

the outcome is also a potential problem in both the candidate gene studies and GWAS. Preeclampsia diagnosis is based on both blood pressure and protein measurements, as well as syndromic clinical and symptomatic criteria, which can vary in severity over time. Information may be recorded on a prenatal record, through the place of delivery, or in registry data. Because of this, some cases may be missed. In addition, although most of the studies use a similar definition of preeclampsia, some define the phenotype differently, such as including pregnancy induced hypertension without proteinuria or using change in blood pressure rather than an absolute measure as a diagnostic criterion.

One final limitation of the current genetic preeclampsia studies is that very few include child (fetal) DNA in analysis. Although preeclampsia is a maternal disease, it likely originates in the placenta, tissue of fetal origin. Only a few studies (see Table 1) have included fetal DNA, but fetal contribution may explain some of the missing heritability of preeclampsia. In a study by Hill and colleagues, they found only a fetal effect of the variants studied.[194] In another study by Petry and colleagues, in which they solely studied known imprinted genes, they found both maternally and paternally transmitted fetal genes associated with maternal mean arterial blood pressure.[195] Since maternal and fetal genotype are correlated, there may be confounding of the fetal association with preeclampsia by maternal genotype; thus, it is important to include both maternal and fetal genotype in the analysis.

This study contributes to the current state of the genetic preeclampsia research by taking a somewhat hybrid approach. The study is hypothesis driven, as with the candidate gene studies, however, it assesses genes within larger pathways based on biological plausibility of the role of an entire pathway. This reduces the number of SNPs from a GWAS to improve power while still looking at a broader set of variants. In addition, we use evidence of reported associations from phenotypes that are likely related to preeclampsia, such as general hypertension and cardiovascular disease, areas in which the state of the genetic literature is much more advanced. Wan and colleagues did this successfully in two recent papers – they

selected two variants that were associated with hypertension in GWAS, and found they were also associated with preeclampsia.[196,197] By prioritizing the variants with known associations, they could reduce the need for correcting for multiple testing, greatly improving power. At the same time, we account for the contribution of both maternal and child genetics by implementing a dyad analysis design.

D. Knowledge Gaps

Inconsistent genetic findings and novel genes

This dissertation investigates new genes and their association with preeclampsia while building on those associations already suggested. Many of our genes had been studied in relation to cardiovascular disease broadly, but never investigated with preeclampsia. To date, only two specific genes in nitric oxide and carbon monoxide pathways (*NOS2, NOS3)* have been reported on for the study of preeclampsia.[17,82,198,199] Most studies of preeclampsia seeking to look at several genes have used microarray technology to assess differential gene expression patterns in placentae.[123] However, with the exception of genes coding for leptin (*LEP*) and fms-related tyrosine kinase 1 (*FLT1*), these studies have found inconsistent results of differentially expressed genes in preeclampsia.[64,200–210] There have been over five hundred papers of candidate gene studies in preeclampsia, but most have investigated a single polymorphism in a single candidate gene. Although over 50 candidate genes have been studied and reported, eight genes account for approximately 70% of the published research about candidate genes for preeclampsia.[123] *NOS3* is one of the more widely studied genes that we investigate[155,188], but the other genes are novel for this question. Aside from *NOS2* and *NOS3,* genes involved in carbon monoxide and nitric oxide signaling have not been extensively investigated for their relationships with preeclampsia.

Sample size and measurement of preeclampsia

This study uses a large sample of validated cases and validated controls to undertake a mother-child dyad approach. Most of the genetic studies of preeclampsia to date have used relatively small sample sizes. These studies may be underpowered, particularly to address heterogeneity of disease. Another difficulty in the genetic study of preeclampsia is inconsistency in case definition. Although preeclampsia is generally defined as gestational hypertension plus proteinuria, the cut-points for those criteria can sometimes vary by medical provider and have changed over time.[211] This imprecision in phenotype definition may increase random variation in the outcome, decreasing power overall.

Family-based methods to disentangle maternal and child genetic effects

A unique challenge of studying the genetics of preeclampsia is that both maternal and fetal genes may play a role. One cannot determine the maternal and fetal genetic contribution with a traditional case-control design because mother and child genotypes will be correlated. This is a potential concern in many reproductive outcomes, and one that has not yet been thoroughly explored.[212] Very few studies of preeclampsia have investigated the fetal genotype,[123] and none have used both mother and child DNA to control for confounding of fetal genotype by maternal genotype.

The case-parent trio design is one that is sometimes used in studies of rare outcomes such as cancer, because the design significantly improves power, reduces the potential of population stratification bias, and removes the requirement of recruiting controls. A similar design, a mother-child dyad design may similarly improve power with less burden of recruitment. Shi and Weinberg have explored some of these methods,[213] and this study continues to apply their innovative methods that may be so relevant to reproductive epidemiology.

This study uses a mother-child dyad approach. The sample consists of 1564 validated preeclampsia case-pairs and 999 validated non-preeclampsia control-pairs. This innovative approach allows us to disentangle maternal and fetal genetic contributions to preeclampsia by controlling for both maternal and fetal genotype while examining genetic effects.

Investigation of gene by environment interactions, an emerging area in reproductive epidemiology

This study investigates gene by environment interactions, an emerging area in reproductive epidemiology. To my knowledge, the only other study of gene by environment interactions for preeclampsia is one assessing the interaction of the *MTHFR* gene and folate.[191] There have been a few other studies of gene by smoking interactions in the field of reproductive epidemiology on birth defects and preterm birth. One assesses the association of *TNFα*, smoking, and cleft lip and palate,[191] and another of *GSTT1*, smoking, and orofacial clefts.[214] Another is assessing *CYP1A1, GSTT1*, and their interactions with smoking on preterm birth.[215] Although these studies provide some evidence of gene by environment interaction, they do not analyze the unique combination of our selected genes, exposure, and outcome.

The large, inverse association of smoking and preeclampsia has been found repeatedly, but it is still unknown as to whether this association is indeed causal or rather due to some pervasive and uncontrolled bias (e.g. left truncation). Identifying a gene by environment interaction with smoking would provide evidence that the inverse relationship may at least in part be due to an underlying biological mechanism. Since there is such a strong and consistent, yet unexplained relationship with the environmental exposure of smoking and preeclampsia, this study provides an opportunity to investigate this complex relationship further.

Genetic epidemiology studies, particularly GWA studies are moving toward novel methods to assess gene by environment interactions such as gene by environment-wide interaction studies (GEWIS). Much of these are with the goal of characterization of joint effects,

rather than just modeling and testing significance of interaction terms, as has been typical in past studies, as well as discovery of new loci.[216] It is possible that genetic factors that impact disease through environmental interaction will not be detected through traditional GWAS analyses.

The study of interaction of genes in pathways relating to carbon monoxide signaling, nitric oxide signaling, and detoxification with the environmental exposure of cigarette smoking offers an opportunity to explore both a potential statistical and biological interaction, as there is evidence that both endogenous and exogenous CO and NO and well as other cigarette smoke components may affect preeclampsia. Family-based designs are also an area in which G x E methods have just recently been developed,[212,217,218] yet are a promising technique for improving power for G x E studies, which otherwise need large sample sizes.[216] Although this study is a candidate gene study, it is possible that the use of maternal and fetal DNA in dyad methods can be expanded and included with novel methods in GEWIS for reproductive outcomes that likely have very important genetic and environmental risk factors, both through maternal and fetal pathways.

In summary, this study uses a large and validated sample of cases and controls from both mothers and their offspring. The maternal-child dyad approach allows us to investigate contributions of maternal and fetal genetics, rarely done in reproductive epidemiology studies. GWA studies of preeclampsia have also only generally used maternal blood or buccal cells, and have not incorporated fetal genetics into analysis. Dyad analytic methods can improve efficiency in study design and analysis both for genetic studies and those of gene by environment interactions by imposing constraints inherent in the family structure of dyads. Genome wide association studies for preeclampsia have been conducted with inconsistent results; our hypothesis-driven candidate pathway study investigates genes for which there is a plausible biological mechanism yet have not been thoroughly investigated. We seek to provide some

insight into the paradoxical relationship of smoking and preeclampsia, a disease for which there is still no known cause or effective therapeutic treatment.

CHAPTER III. METHODS

A. Study Design Overview

This study is a case-control study of validated preeclampsia cases and non-preeclampsia controls nested within the Norwegian Mother and Child Cohort Study (MoBa), conducted by the Norwegian Institute of Public Health.[219] Preeclampsia is a rare disease, occurring in approximately 4% of pregnancies in our study population; therefore, a case-control study is the most efficient way to study this outcome. A candidate gene case-control study allows us to look for gene by smoking interaction. To elucidate whether any effects of genetic variants on risk of preeclampsia are of maternal or fetal origin, we use a mother-child dyad design; cord blood samples have been analyzed from each infant born to the case and control mothers if available. A candidate pathway case-control study best enables us to efficiently determine if genetic variants in the proposed pathway are associated with preeclampsia, whether these variants in the mother or child have a greater effect, and if they are modified by maternal smoking during pregnancy.

B. Study Population

Norwegian Mother and Child Cohort Study

The Norwegian Mother and Child Cohort Study (MoBa) is a large prospective birth cohort of pregnant women and their offspring, recruited throughout Norway from 1999 to 2008. The purpose of the study is to identify causes of disease in the mother and child.[220] All pregnant women living in Norway who gave birth at a hospital or maternity unit with more than 100 births

annually and could speak Norwegian were eligible; MoBa investigators applied no other exclusion criteria. Pregnant women were recruited by mail with addresses provided by their prenatal care provider prior to their routine ultrasound appointment scheduled for 17 to 20 weeks' of gestation. Because of the consistency and availability of care in Norway, nearly all pregnant women have routine ultrasounds at this time.[221] Of all women invited to participate, 38.5% enrolled in the study.[219] Participants complete several questionnaires by mail. Pregnant women were recruited by mail prior to their routine ultrasound appointment at 17 to 20 weeks' of gestation. Of all women invited to participate, 41% enrolled in the study.[222] Participants completed two prenatal questionnaires about their health and environment. The early pregnancy questionnaire is completed prior to the ultrasound appointment at 13-17 weeks' gestation. It asks for detailed information about previous pregnancies, medical history, medications, occupation, home and work exposures, mental health, and lifestyle habits. A late pregnancy questionnaire (~30 weeks' gestation) asks about the pregnancy and any changes in status from the first questionnaire. Survey completion rate was 91% for the early pregnancy questionnaire (administered between weeks 13 and 17) and 83% for the late pregnancy questionnaire (administered at week 30).[222] Biospecimens are also collected from the mother and child. For those enrolled, maternal blood was collected at the ultrasound appointment. Cord blood was collected from the child at birth unless they were unable; in those cases, blood was collected by heel stick during routine PKU screening 3 or 4 days after birth. Maternal blood was received from 89% of participants and child (cord) blood from 81% of children in the cohort.[222] For both maternal and infant blood, DNA was extracted at the time of collection before being stored at the MoBa Biobank.

A validation study of preeclampsia diagnosis was conducted within the MoBa cohort.[223] The investigators selected all pregnancies with preeclampsia in MoBa that were registered in the medical birth registry of Norway (MBRN) (N=4081), and a random control group without preeclampsia registered in the MBRN (N=2000). Delivery units were asked to provide antenatal

charts that contained blood pressure and urinary measurements, as well as hospital discharge codes. Data from 87% of eligible pregnancies was received (N=5340). The investigators considered a gold-standard for true preeclampsia to be blood pressure of 140 mmHg systolic and/or 90 mmHg diastolic after the 20th week of gestation, together with proteinuria of $\geq$ 0.3 g per 24 hours ($\geq$1 + on dip-stick) noted on the antenatal chart, or presence of preeclampsia/eclampsia ICD-10 codes on the hospital discharge form. One antenatal visit that fulfilled these criteria was considered adequate to be a preeclampsia case. Of the 3500 registered preeclampsia cases and 1840 registered to be unaffected by preeclampsia for which records were received, 2936 pregnancies identified as preeclampsia cases from the MBRN were verified to have been affected by preeclampsia, and 1745 pregnancies identified as unaffected by preeclampsia were found to be negative for preeclampsia. Overall positive predictive value considering registry data as predictive of clinical records in the validation study was found to be 83.9%. When extrapolated to the entire MoBa population, estimated sensitivity was low (43.0%) and specificity was high (99.2%), and many false negative cases in MoBa actually had mild forms of preeclampsia. [223]

The samples for the proposed study were selected only from the preeclampsia cases and non-preeclampsia controls validated within the study described above. When controls were determined to meet case criteria in the validation study, they were reassigned to the case group.

Eligibility and Sampling

As described above, for this proposed study, 2936 cases were selected from women with validated preeclampsia and 1745 controls were selected from women who were reported not to have preeclampsia and were validated as non-cases. Cases were oversampled relative to controls in order to maximize power to detect associations in subgroups of PE (e.g. early PE, severe PE). Because of inconsistencies with diagnosis of preeclampsia,[211] all cases and

controls were validated through an independent validation study described above.[223] Women with pregnancy-induced hypertension (~1% of sample) were not excluded from controls. To be eligible for this study, women also met the following criteria:

1) Singleton pregnancy conceived without the use of in vitro fertilization who returned both the 1st and 3rd pregnancy questionnaire (which contains self-reported smoking information)
2) Maternal blood sample collected during mid-pregnancy and DNA extracted and stored in the MoBa Biobank
3) No evidence of hypertension prior to pregnancy

In total, 2682 preeclampsia case samples (2236 samples from 1118 mother/child pairs and 446 unpaired maternal samples) and 1967 non-preeclampsia control samples (1936 samples from 968 mother/child pairs and 31 unpaired maternal samples) met inclusion criteria and were genotyped. There were no unpaired offspring genotyped.

Study Sample for Aim 1

Quality control measures are described in the section on exposure assessment. After quality control, the final analysis sample for aim 1 (n=4551 total samples) included dyads with both mother and child genotype data as well as incomplete dyads with only mother or child genotype data (Table 2).

Table 2. Final analysis sample for aim 1.

	Cases			Controls			Total		
	Mother	Child	Total	Mother	Child	Total	Mother	Child	Total
Paired Dyad	1076	1076	2152	935	935	1870	2011	2011	4022
No paired DNA	459	10	469	46	14	60	505	24	529
Total	1535	1086	2621	981	949	1930	2516	2035	4551

Study Sample for Aim 2

The sample for aim 2 made additional exclusions for ancestry as described in paper 2. After quality control, the final analysis sample for aim 2 (n=4514 total samples) included dyads with both mother and child genotype data as well as incomplete dyads with only mother or child genotype data after excluding samples in which the first three principal components were greater than 3 standard deviations from the mean (Table 3).

Table 3. Final analysis sample for aim 2.

	Cases			Controls			Total		
	Mother	Child	Total	Mother	Child	Total	Mother	Child	Total
Paired Dyad	1063	1063	2126	925	925	1850	1988	1988	3976
No paired DNA	450	20	470	45	23	68	495	43	583
Total	1513	1083	2596	970	948	1918	2483	2031	4514

Human Subjects

Women provided informed consent prior to participation in the Norwegian Mother and Child Cohort Study. MoBa received approval from the Institutional Review Board of the Norwegian Institute of Public Health. The parent study in which the samples were genotyped was approved by the Institutional Review Boards of the Norwegian Institute for Public Health and the University of North Carolina at Chapel Hill.

This dissertation uses previously collected, de-identified data and no study participants were contacted for additional information. Data were encrypted and password protected. The project was approved by the Institutional Review Board of the University of North Carolina at Chapel Hill on April 5, 2015.

C. Outcome Assessment

Details of the validation study for preeclampsia have been previously described.[223] In brief, preeclampsia was defined by the American College of Obstetrics and Gynecologists (ACOG),[27] which includes the following two criteria:

1) Systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure of ≥ 90 mm Hg occurring after 20 weeks' gestation in a woman whose blood pressure has been previously normal, and

2) Proteinuria, with excretion of ≥ 0.3 g of protein in a 24-hour urine specimen or as measured by 1+ on urine dipstick.

This 2013 revision of this definition included clinical symptoms, however, we used the criteria current at the time of the validation study. Diagnoses of preeclampsia with severe features, eclampsia, or HELLP syndrome (a variant of preeclampsia marked by hemolysis, elevated liver enzymes, and low platelet count) are often considered indications for induction of delivery. Thus, these cases often go into the hospital system immediately after diagnosis and information cannot be validated solely from the antenatal record. We therefore considered as "true cases" any case with an ICD-10 code of severe PE (O14.1) or HELLP syndrome (O14.2) and delivery <39 weeks. All cases with an ICD-10 code of eclampsia are routinely validated by the MBRN and are included as true cases.

We also performed analyses within subtypes of preeclampsia. Although ACOG guidelines also include clinical symptoms for the diagnosis of severe preeclampsia, study data were limited to blood pressure and urinalysis values recorded on the antenatal chart, along with birth outcome features captured by the Medical Birth Registry. Criteria for these subtypes are as follows:

Severe preeclampsia

In this study, severe preeclampsia meets the general criteria described above and at least one of the following additional requirements in the antenatal record[27]:

1) Systolic blood pressure of at least 160 mm Hg or diastolic blood pressure of 110 mm Hg, or

2) Urinary protein excretion of at least 5g/day in a 24-hr urine specimen or at least 3+ urine dipstick protein measurement.

Additionally, cases in this study were classified as severe if they were diagnosed as having eclampsia or HELLP syndrome in the MBRN.

Early-onset preeclampsia

Early-onset preeclampsia is preeclampsia that meets the any criteria for preeclampsia diagnosed prior to 34 completed weeks of gestation.[27] Estimated gestational age at time of preeclampsia diagnosis was recorded in the antenatal record. Cases were also considered early-onset if they met the criteria for preeclampsia as described above and had a delivery before 34 weeks' gestation.

Preeclampsia with delivery prior to 34 weeks' gestation

The subset of early-onset preeclampsia cases that also had a delivery before 34 weeks' gestation were also considered separately.

Preeclampsia complicated by small for gestational age

Preeclampsia complicated by small for gestational age was defined as cases meeting the criteria for general preeclampsia described above plus an infant born small for gestational age (<10th percentile). Gestational age provided from ultrasound estimation and birthweight were attained from the MBRN. Norwegian population-based tables for fetal term prediction and size assessment were provided from eSnurra Norway (http://www.nsfm.no/esnurra/0.php) to compare percent deviation from median birthweight for each gestational age and with the population percentiles to determine whether an infant falls below the 10th percentile.

D. Exposure Assessment

Sample Collection and DNA Extraction

Maternal and child blood samples were collected for the MoBa study as described above. For DNA extraction, whole blood was collected in a 7-ml EDTA tube (Becton-Dickson, Plymouth, UK), labeled with the woman's name and national identification number, and shipped overnight to the MoBa Biobank, where samples were processed on the day they were received. Once received at the Biobank, whole blood was aliquoted into two polypropylene deep-well plates (ABgene, Surrey, UK) using Tecan pipetting robots. DNA was then extracted manually using the FlexiGene kit (Qiagen, Hilden, Germany). Quality control was performed on all DNA samples using a spectrophotometer (Spectramax 190, Molecular Devices, Sunnyvale, CA) to test optical density. DNA needed a purity of 1.6-2.0 260/280 ratio, a concentration greater than 20ng/µl, and an optical density at zero for the negative control to be included in each 24-sample batch. DNA was then aliquoted into 1.4-ml deep-well plates and stored at -20 °C, where connected to backup power and an alarm system. Samples were collected and processed at the hospital before being sent to the MoBa Biobank.[221]

Illumina Process

The Illumina HumanCoreExome+ assay (Illumina, Inc., San Diego, CA) has been described in more detail by Illumina.[224] The HumanCoreExome+ BeadChip includes approximately 240,000 genome wide association markers and 240,000 exome markers plus space for additional custom markers.

Tag Selection for Parent Study

Custom selected SNPs included SNPs on the Illumina Cardio-Metabolic chip not already included in the HumanCoreExome+ manifest, with particular emphasis on three regions of interest: 1) regions associated with systolic blood pressure, diastolic blood pressure, and hypertension; 2) regions associated with myocardial infarction, chronic heart disease, and chronic kidney disease; and 3) regions associated with body mass index, lipids, and C-reactive protein. Additional candidate genes were selected based on the following pathways and sources: 1) existing associations with preeclampsia and/or commonly hypothesized genes, 2) inflammation, 3) angiogenesis, 4) apoptosis, 5) smoking detoxification, 6) carbon monoxide signaling, 7) smoking addiction, and 8) novel pathways including Vitamin D and in vitro studies. For each gene, the Illumina database was queried for all polymorphism design scores within our genes of interest, allowing for 20kb upstream and 10kb downstream margins. A scoring algorithm for each SNP was created, taking into account Illumina design score, Illumina error codes, DNA coding changes, and presence in a possible 5' promoter site. The composite SNP database was then analyzed using TagZilla (http://tagzilla.nci.nih.gov) to identify haplotype tagging SNPs with an R^2 criteria of 80%. In total, 525,125 variants were genotyped for the larger parent study of genetics and preeclampsia.

Gene and SNP Selection

Gene and SNP Selection for Aim 1

For aim 1, 66 genes involved in three canonical pathways were selected for analysis from the parent study. Genes were selected from the following canonical pathways relevant to carbon monoxide and nitric oxide activity: 1) endothelial nitric oxide synthase (eNOS) signaling, 2) heme degradation, and 3) hypoxia-inducible factor 1-alpha (HIF1A) signaling. (Table S1). Particular emphasis was placed on genes in which there were prior associations with other cardiovascular disease or are hypothesized to be involved in reproductive processes

SNPs were extracted for this analysis from the overall study database after genetic quality control using PLINK 1.07. SNPs were extracted for each gene based on position of 10kb upstream and downstream margins around the transcription start and end sites for each gene. After QC, 1518 SNPs were selected and analyzed for aim 1.

Gene and SNP Selection for Aim 2

For aim 2, 124 genes involved in response to cigarette smoke components were identified from 8 canonical pathways. These included the same pathways in aim 1: 1) endothelial nitric oxide synthase (eNOS) signaling pathway, 2) heme degradation, 3) hypoxia-inducible factor 1-alpha (HIF1A). They also included five additional pathways related to smoking detoxification: 4) xenobiotic metabolism, 5) aryl hydrocarbon receptor signaling, 6) glutathione-mediated detoxification, 7) nicotine degradation II, and 8) nicotine degradation III. These pathways and their associated genes are described in Table S2. A total of 1,915 SNPs were selected and analyzed for aim 2, using a 10kb upstream and downstream margin around the transcription start and end sites of each gene.

Genotyping and Quality Control

SNPs were genotyped by the UNC Mammalian Genotyping Core using the HumanCoreExome+ chip from Illumina (Illumina, Inc., San Diego, CA). Thirty-five samples failed initial genotyping at the lab due to chromosomal aberration (1), DNA contamination (14) or low concentration, no DNA, or poor quality DNA (20). These samples were not included in the QC process. Genotyping of 12 samples on one plate (NPE-617) was repeated due to a defective chip; failed samples were dropped and repeated samples were retained for analysis. A total of 525,125 variants were typed.

This study is not powered to assess extremely rare variants, and SNPs with a minor allele frequency of <5% were excluded. 118,391 variants had a minor allele frequency of <5% and were excluded from analysis.

Hardy Weinberg Equilibrium (HWE) was examined in PLINK among non-cases stratified by relationship type (mother or child) ($p<0.001$). A QQ plot of observed versus expected Hardy-Weinberg p-values was generated to determine an appropriate cut point for deviation. SNPs were determined to be outside of HWE if $p<0.001$. Any SNPs showing evidence of Hardy Weinberg disequilibrium were examined by Jason Luo of the Mammalian Genotyping Core to determine possible reasons and if disequilibrium was due to genotyping error. SNPs that showed HWE disequilibrium after review were dropped from analysis (n=9,457).

Data were cleaned using best practices as described by the Cohorts in Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium.[225] Data were cleaned using PLINK 1.07. (http://pngu.mgh.harvard.edu/purcell/plink). Monomorphic loci were dropped (n=153,659). Both subjects and SNPs where call rates were <95% were dropped. All samples had genotyping information for at least 95% of genotypes and none were excluded due to missing genotype information. 123 SNPs were missing genotype information for all individuals and were excluded. One additional SNP (exm2216494) was missing genotype information for >95% of individuals and was excluded. In total, 124 variants were excluded due to missing information for at least 95% of individuals.

Sex-specific markers were genotyped and inspected. Three individuals were excluded because reported sex phenotype was female and sex based on heterozygosity was determined to be male. Inbreeding coefficients were calculated using PLINK 1.07 (plink --file data --het) based on observed number of homozygous genotypes versus expected number of homozygous genotypes. An exclusion threshold of $F \leq -0.15$ was used. No individuals met this threshold and were excluded from the analysis based on inbreeding coefficient.

The study sample uses data from a limited geographic area and an ethnically homogeneous population which presents the possibility of individuals to be related. Statistical analyses make the assumption of independence of samples, and residual relatedness can lead to an overestimation of the effect of a variant on the outcome because genotypes in families will be overrepresented. Heritability was examined using PLINK identity-by-descent (IBD) to confirm parent-child relatedness, and to discover any fugitive relatedness in the cohort. Cryptic associations were assessed in PLINK 1.07 using identity-by-descent analysis. PLINK provides an estimate of proportion IBD (pi-hat) between each pair of samples. Identical samples (or monozygotic twins) have an expected pi-hat=1, parent-offspring or sibling pairs have an expected pi-hat=0.5, second-degree relatives (aunt-niece/nephew) have an expected pi-hat =0.25, and third-degree relatives (first cousins) have an expected pi-hat=0.125
Pairs with a high degree of relatedness (pi-hat>0.125) were flagged and inspected. Pairs which were indicated to be related in MoBa but were not confirmed to be related by IBD analysis were dropped. Pedigrees were examined to retain the most complete data among families. In instances where there were two related mothers (sisters, cousins) the mother with associated child data was retained. If both mothers had paired child data, then the pair with the least amount of missing SNP data was retained. In instances in which there were two related children for a single mother (siblings or half-siblings), the closest relationship was selected, and otherwise one sibling was randomly selected.

One pair in which one of the samples was a mother and one was a child was unexpectedly found to be a duplicate. This pair of samples was dropped since one was likely a sample swap. Among pairs that were not labeled as being related, 55 pairs were found to be genetically related at a pi-hat>0.125. Some of these pairs were part of the same pedigree, so ultimately 42 separate samples were excluded because of relatedness.

IBD analysis in PLINK was also used to confirm relatedness in mother-child pairs that were labeled as being related. Mother-child pairs should have a pi-hat value of approximately

0.5. Of labeled related pairs, 9 were found to be genetically unrelated. In these cases, both mother and child were dropped, as there is no certainty which sample is mislabeled.

Several measures were put in place in this study to ensure quality control of genotyping. Since dyad analysis is dependent on paired DNA, linked mother and child DNA were analyzed on the same plate to reduce the possibility of losing information from both if a problem occurs, but cases and controls were randomly distributed by plates to reduce systematic error.

Blind duplicate samples were included to identify genotyping errors by the examination of discordant alleles between duplicate samples. For this analysis, both study (MoBa) duplicates and CEPH (Centre d'Etude du Polymorphisme Humain, a known family trio) controls were included. In total, 186 unique samples were a part of the blind assessment.

Fifty-one duplicate pairs were chosen randomly from the study sample to be included as duplicates (Table S3). Duplicate samples with high levels of discordance (>2 instances of SNP mismatch) were excluded. Two samples were excluded due duplicate mismatch (Table S4). After exclusion of discordant samples, all SNPs were highly concordant among duplicate samples, and no SNPs were excluded for analysis based on discordance.

CEPH duplicates (n=208 pairs), representing five families composed of 15 individuals from the Coriell Institute CEPH Utah pedigrees were selected and genotyped (Table S5). Twenty-eight trios of 84 samples were included for quality control. Each sample was repeated between 4 and 9 times over the course of the assay. Between one and seven CEPH controls were included on each plate (Table S6). Ten of these CEPH duplicate samples were included on the same plate and 74 were included on different plates, which allowed for inter- and intra-plate quality control (Table S7).

Out of the 208 CEPH blinds, there were three pairs of duplicate samples that each had two discordant SNPs. The two discordant SNPs were the same for each duplicate pair (rs7719740 on chromosome 5 and rs4820872 on chromosome 22). Although the same two SNPs were discordant in three separate pairs, the discordance was driven by one sample that

was repeated in each pair; the other duplicates of this set were all 100% concordant. (Table S8). Therefore, these two SNPs were not dropped from the analysis. All other CEPH duplicates had identical genotypes. These error rates were within our pre-specified range of acceptable values.

Known CEPH trios of child, mother, and father were also included for quality control by confirming mendelian inheritance. Five families composed of 15 individuals from the Coriell Institute CEPH Utah pedigrees were selected and genotyped (Table S5). Twenty-eight trios were included for quality control.

PLINK version 1.07 was used to assess mendelian inheritance among the 28 trios (plink --bfile cephtrios --mendel). Overall genotyping was 0.999237 and 547,082 variants that passed initial quality control screening were included in this step. No mendelian errors were detected. No SNPs were excluded from analysis based on quality control assessment of CEPH trios.

Quality Control Summary and Flowcharts

Individual Samples

Genotyping was conducted on an Illumina HumanCore ExomeChip+ platform with 4,799 samples typed for 545,125 SNPs. The samples included women for this study, quality control samples, and repeated samples due to initial genotyping error.

Of these 4,799 samples genotyped, the following were removed during the QC process:

84 CEPH trio samples included for consistency of duplicates and mendelian inheritance QC
22 known duplicates included for QC
12 repeated samples due to defective chip
3 repeated samples due to lab error
35 that failed initial genotyping for other reasons:
    1 due to chromosomal aberration

14 due to DNA contamination

20 due to low concentration, no DNA, or poor quality DNA

4,643 samples were assessed in PLINK for further QC.

92 samples were excluded during this phase:

29 blind duplicates included for QC

2 samples discordant with blind duplicates

3 sex discrepancies

58 related subjects with >0.125 IBD

No individuals had >5% missing genotype data or inbreeding problems (F<-0.15 inbreeding coefficient).

In total, 248 individuals were dropped from the original genotyped samples and 4,551 were available for analysis. Table S9 provides IDs and reasons for all sample exclusions made during the quality control process.

Single Nucleotide Polymorphisms

Samples were typed on the HumanCore ExomeChip+ platform for 525,125 total SNPs. 123 SNPs were missing genotype information for all individuals and were excluded (Table S6). One additional SNP (exm2216494) was missing genotype information for >5% of individuals and was excluded (Table S10).

545,001 SNPs entered the QC process. Of these:

153,659 were monomorphic.

9,457 were outside of HWE p<0.001

118,391 had a minor allele frequency of <5%

In total, 281,631 SNPs were dropped from the original typed variants and 243,494 SNPs were available for analysis.

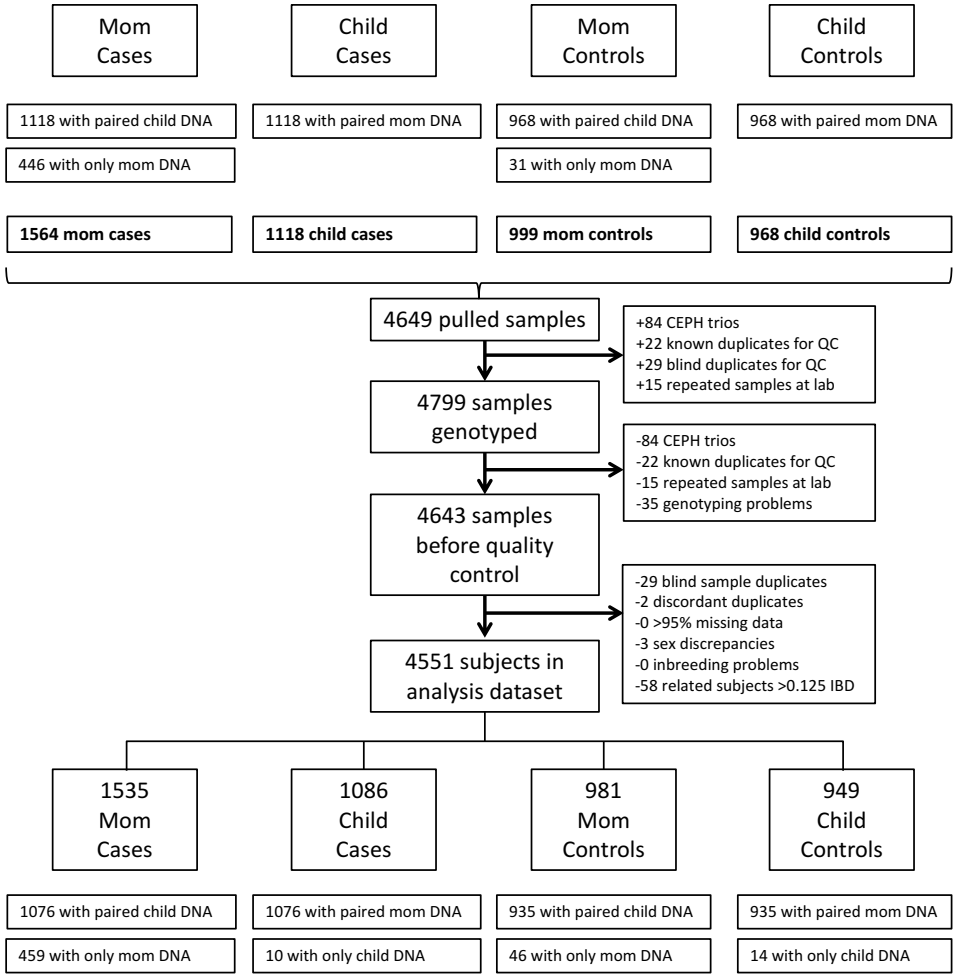


Figure 7. Flowchart for sample quality control process.

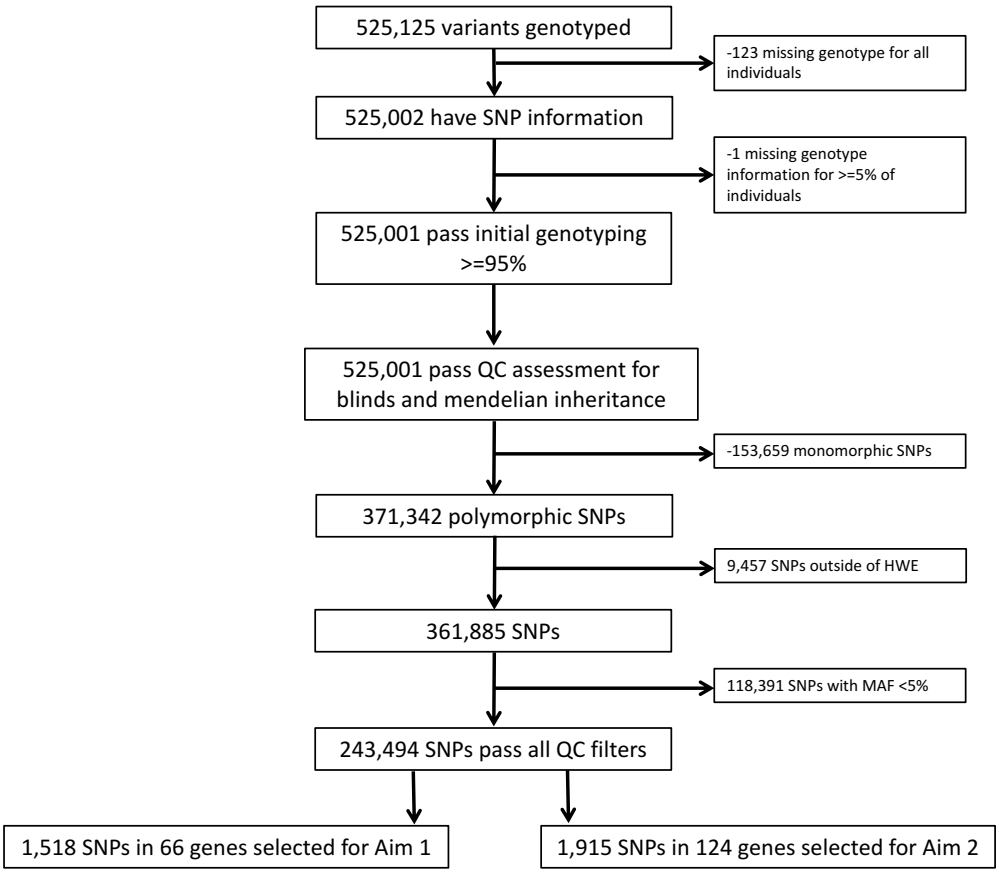


Figure 8. Flowchart for SNP quality control process.

Genetic Ancestry

As indicated the directed acyclic graphic description for aim 1, ancestry may confound the relationship between a genetic variant and the outcome, known as population stratification bias. In genetic studies that employ logistic regression, eigenstrata are generally included as covariates to reduce population stratification bias, however, the use of log-linear Poisson regression with a mating type parameter in the model should reduce the need for adjustment with eigenstrata. Nevertheless, because of concern that missingness may depend on membership in subpopulations, we assessed ancestral origin in our population.

Quantile-quantile plots and calculation of genomic control lambda[226] were assessed to detect evidence of population stratification and indicated no systematic test statistic inflation, suggesting that population stratification was negligible (Figure 9).

Figure 9a. Figure 9b.

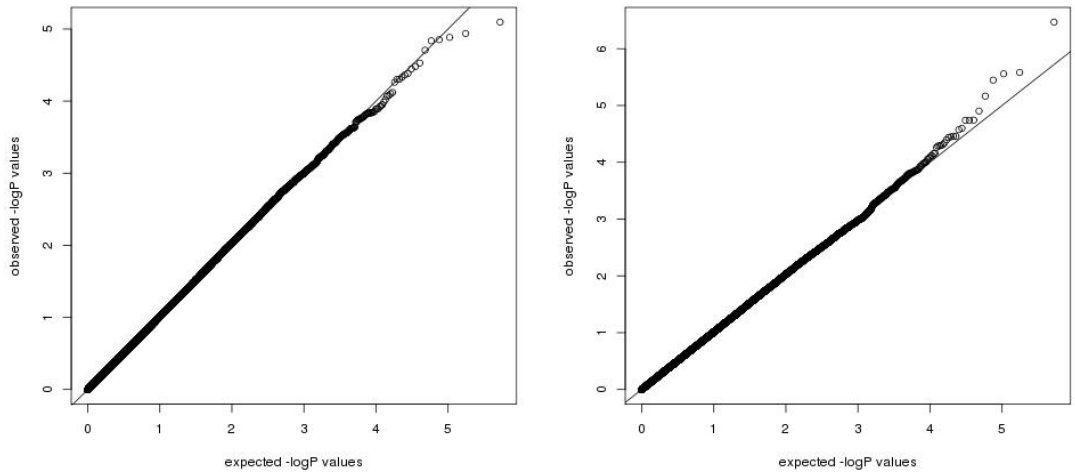


Figure 9. Quality control quantile-quantile plots for maternal and child genotypic effects among genome-wide data of 263,494 variants to assess genomic inflation. Test results (observed -$\log_{10}$p values) are plotted against the expected -$\log_{10}$p values for each of 263,494 SNPs in the sample. Figure 9a. Q-Q plot of maternal genotypic effects among maternal samples. Genomic inflation factor, lambda=1.01. Figure 9b. Q-Q plot of child genotypic effects among child samples. Genomic inflation factor, lambda=1.03.

The top 3 principal components of genetic variation were plotted for the MoBa data together with the 1000 Genomes reference populations and visually inspected to assess evidence of admixture (Figure 10).

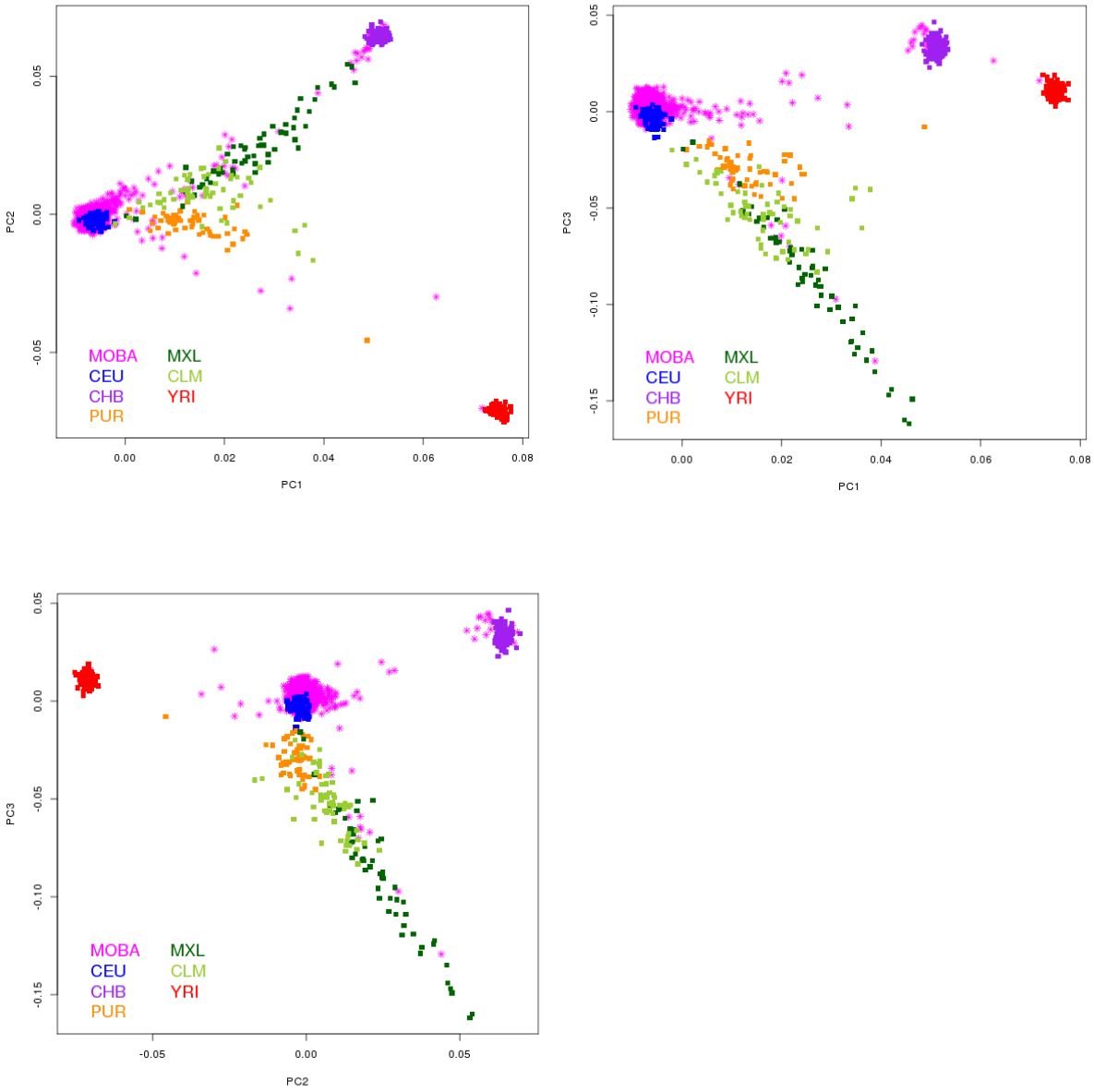


Figure 10. Top three axes of genetic variation based on common SNPs for our Norwegian Mother and Child Cohort sample (MOBA) compared to 1000 Genomes reference populations. Plots are show for axes 1 and 2, axes 1 and 3, and axes 2 and 3. Reference populations are: CEU: Utah Residents with Northern and Western Ancestry; CHB: Han Chinese in Beijing, China; PUR: Puerto Ricans from Puerto Rico; MXL: Mexican Ancestry from Los Angeles, USA; CLM: Colombians from Medellin, Colombia; YRI: Yoruban in Ibadan, Nigeria.

Cases and controls were also plotted along ancestral axes to examine differential case and control status by ancestry (Figure 11).

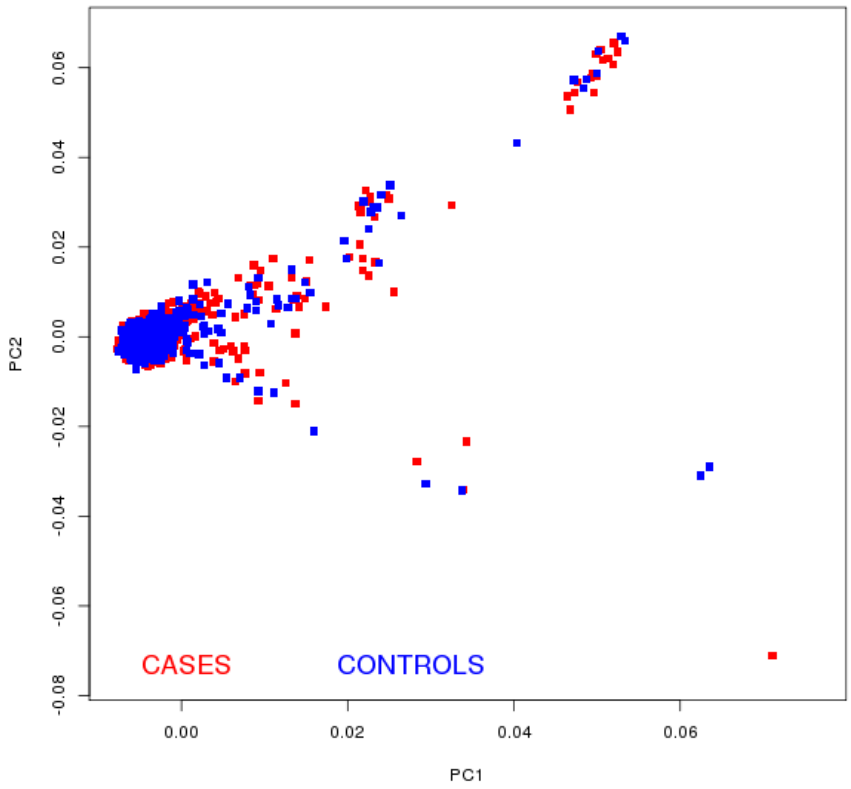


Figure 11. Case and control status of the study sample plotted along the first two axes of variation for genetic ancestry.

Samples that fell outside of the European ancestral population were excluded in a sensitivity analyses for Aim 1. For aim 2, we excluded all samples that fell outside 3 standard deviations from the mean for the top three principal components.

Maternal Smoking

Maternal smoking was included an effect measure modifier in our study to assess gene by environment interactions. Smoking is self-reported on the early and late pregnancy MoBa questionnaires through a series of questions that ask about current smoking, smoking prior to

pregnancy, and smoking in the intervening window between questionnaires; quantity of cigarettes smoked; age at smoking initiation; whether the woman stopped smoking; age or week of pregnancy when stopped smoking; and time of day when smoking primarily occurs. Smoking was examined at three time periods, as these provide different and important information. Smoking during the first trimester will always precede preeclampsia, while symptoms of preeclampsia may exist before smoking is reported on the third questionnaire. Smoking as late as the third trimester, however, likely indicates both a greater level of smoking and a greater addiction, as women are encouraged to quit as soon as they become pregnant. As with all self-reported smoking variables, there is potential for underreporting. However, a recent validation study of self-reported smoking and plasma cotinine indicated that reported smoking on the MoBa questionnaire is a valid marker of tobacco exposure (Sensitivity=82%, Specificity=99%).[227] Women who falsely reported non-smoking compared to women who correctly reported non-smoking were more often cohabiting (61% vs 42%) rather than married (35% vs 56%) and had lower education (55% vs 37% with education 12 years or fewer).[227] Self-reported maternal smoking in the MoBa cohort is 23% in the first trimester and 9% in the third trimester. Self-reported maternal smoking in the study sample was 19% in the first trimester and 8% in the third trimester.

We created a dichotomous time-specific smoking variable for any maternal smoking during the window of 11 to 20 weeks' gestation. We used smoking during this time period for our primary analysis, as we considered it to be the period of greatest biological relevance to our hypothesis during which trophoblast invasion, spiral artery remodeling, and maternal blood flow perfusion into the intervillous space occur.[228] If a person indicated she had never smoked, we considered her a non-smoker at our time point. If she indicated she did not currently smoke at the time of the early survey and had no other evidence of smoking at any time point, we considered her a non-smoker. If a woman indicated she currently smoked "sometimes" or "daily" in the early pregnancy questionnaire, we considered her to be a smoker. If she indicated that

she did not currently smoke at the time of the survey, but noted a quit week after gestational week 10 on either the early or late questionnaire, we considered her to be a smoker. If a woman was missing smoking information on the early survey, we supplemented with information from the late survey. If she indicated a smoking quit week on the third survey after 10 gestational weeks', we considered her a smoker. If she indicated being a daily smoker in the late survey, we assumed she had been smoking in weeks 11 to 20. Smoking was noted as missing if we were missing information from both the early and late surveys, or smoking status was ambiguous based on the existing information.

E. Additional Covariates

Potential risk factors for preeclampsia include maternal age (both young and old), race, nulliparity, change in partner, long interpregnancy interval, high body mass index, prior preeclampsia, multi-fetal pregnancy, diabetes, and low socioeconomic status. Covariates of interest were obtained from the early pregnancy MoBa questionnaire and the Medical Birth Registry of Norway. The following covariates were considered:

Maternal Age

Maternal age at delivery was recorded on the MoBa questionnaire and in the MBRN.

Parity

Number of previous deliveries obtained from the MBRN. Parity was considered the number of deliveries stated by the mother or the number of previous deliveries reported by the MBRN, beginning at gestational week 16 (1967-2001) or gestational week 12 (2002 to present); we used whichever number reported was greater.

Body Mass Index

Pre-pregnancy body mass index was calculated from self-reported weight and height from early MoBa questionnaire. Implausible values were excluded (n=5).

Maternal Education

Maternal education information was obtained from the questionnaire item asking participants “What education do you and the baby’s father have?” Levels of education included 9-year secondary school, 1-2 year high school, technical high school, 3-year high school for general studies/junior college, regional technical college/4-year university degree (Bachelor’s, teaching, nursing, etc.), more than 4-year university or technical college (Master’s, medical doctor, PhD, etc.), and other education. Participants were asked whether each level was completed or on-going.

Birth Outcomes

Other birth outcome information including gestational age and birthweight were obtained from the MBRN. Small for gestational age was calculated using Norwegian population percentiles. Gestational age and birthweight were assessed as independent covariates and were also used to determine preeclampsia subtypes.

F. Statistical Analysis

Descriptive Statistics

Descriptive statistics were performed for potential covariates. Continuous variables were examined for measures of central tendency and spread, including the mean, median, standard deviation, and interquartile range. Deviations from a normal distribution including presence of outliers, skew, and kurtosis. For categorical or binary covariates, minimum and maximum

values, and spread of data within categories were assessed. Missing covariate values were tabulated for all variables and analyzed for their associations with the exposure and outcome to determine if missing data imputation was to reduce bias due to missingness. Bivariate associations between exposures (smoking) and outcomes, covariates and outcomes, and exposures and covariates were analyzed to inform model-building.

Specific Aim 1 Statistical Analysis

Aim 1 uses log-linear Poisson regression to estimate relative risks for the main effects of CO and NO related genes in the eNOS signaling, heme degradation, and HIF1A signaling pathways on preeclampsia.

Although preeclampsia is a maternal disease, it involves the placenta, an organ of fetal origin. Because of this unique connection, it is plausible that both maternal genotype and fetal genotype may independently or jointly play a role in the development of preeclampsia. Wilson et al. posit that both maternal and fetal genes related to vascular remodeling may be related to preeclampsia risk,[229] and others have recently recognized the importance of investigating both maternal and fetal genotype in obstetric complications.[124,230] Additionally, maternal and fetal genotype are correlated, as half of child alleles are transmitted from the mother, and thus, fetal genotype may confound the association between maternal genotype and outcome, and vice versa.[213,231] Both of these aspects require that maternal and fetal genotype be simultaneously considered as predictors in a single model using either of the proposed methods.

Logistic regression for the estimation of odds ratios is a commonly used analysis technique for case-control studies of candidate genes. We initially fit separate logistic models for maternal and fetal genotype to calculate unadjusted odds ratios to aid in determining if models are behaving as expected, however, other models can be used to incorporate the related mother-child genotype. A logistic model for the association of each SNP with preeclampsia can

be used including parameters for both maternal and fetal genotype. This model is of the following form proposed by Shi et al.[213]:

$$ln\left(\frac{\Pr\,(D|M,C)}{1-\Pr\,(D|M,C)}\right)=\mu+\alpha_1 I_{(M=1)}+\alpha_2 I_{(M=2)}+\beta_1 I_{(C=1)}+\beta_2 I_{(C=2)}$$

where M and C are the number of copies of the variant allele carried by the mother and child, respectively; I is an indicator function which is 1 when the expression is true and 0 otherwise; α_1 and α_2 are natural logs of the mother's risk (S_1 and S_2) when the mother has 1 or 2 copies of the variant allele, respectively; β_1 and β_2 are natural logs of the child's risk (R_1 and R_2) when the mother has 1 or 2 copies of the variant allele, respectively; and μ is the natural log of the underlying risk in the source population.

Logistic regression is a valid technique provided there is no underlying population substructure,[231] however, it does not take advantage of particular features of mother-child dyad case-control data. With mother-child pairs for cases and controls, one can make certain assumptions and impose certain nonlinear constraints that are particular to family data (dyads and triads) to significantly improve power.[213,231,232] This study uses log-linear Poisson regression with the expectation-maximization algorithm as proposed by Shi et al.[213] to enforce these constraints and estimate risk ratios.

As with the logistic model, we must assume the disease is rare in the population and that the population is risk-homogenous and does not covary with allele prevalence across subpopulations.[213] The parent-child relationship, however, also implies certain family based constraints. Additional assumptions can also improve precision:

1) Mendelian inheritance
2) Parental mating symmetry for the studied locus in the source population
3) Allelic exchangeability

As above, M, F, and C are 0, 1, or 2, for the number of variant alleles that the mother, father, or child carries, respectively. Assuming Mendelian inheritance, the expected frequency of parent pairs is denoted by μ_{mf}, in which the mother has m copies of the variant allele and the father has f copies. Triads of mothers, fathers, and children are classified by the number of variant alleles carried by each of the mother, father, and child, which results in a 15-cell multinomial distribution.[212,233,234] With mother-child dyads, this number is reduced to 7 cells by collapsing over the missing fathers. These frequencies are shown in Tables 4 and 5. Mendelian inheritance imposes the constraint that the expected counts for (1, 0) and for (1, 2) sum to the expected count for (1, 1).

Table 4. Expected frequencies of control mother-child pairs under Mendelian transmission of parental alleles.

	C = 0	**C = 1**	**C = 2**
M = 0	$\mu_{00} + (1/2)\ \mu_{01}$	$(1/2)\mu_{01} + \mu_{02}$	0
M = 1	$(1/2)\mu_{10} + (1/4)\ \mu_{11}$	$(1/2)[\mu_{10} + \mu_{11} + \mu_{12}]$	$(1/4)\mu_{11} + (1/2)\mu_{12}$
M = 2	0	$\mu_{20} + (1/2)\mu_{21}$	$\mu_{22} + (1/2)\mu_{21}$

Table 5. Expected frequencies of case mother-child pairs under a multiplicative model for risk.

	C = 0	**C = 1**	**C = 2**
M = 0	$B[\mu_{00} + (1/2)\ \mu_{01}]$	$BR_1[(1/2)\mu_{01} + \mu_{02}]$	0
M = 1	$BS_1[(1/2)\mu_{10} + (1/4)\ \mu_{11}]$	$(1/2)BR_1S_1[\mu_{10} + \mu_{11} + \mu_{12}]$	$BR_2S_1[(1/4)\mu_{11}+(1/2)\mu_{12}]$
M = 2	0	$BR_1S_2[\mu_{20} + (1/2)\mu_{21}]$	$BR_2S_2[\mu_{22} + (1/2)\mu_{21}]$

In addition to Mendelian inheritance, one can also assume parental mating symmetry, that is $\mu_{mf} = \mu_{fm}$ at the locus under study. This assumption implies an additional constraint that the expected difference in the source populations (Table 4) between the count for (1,0) and (0,1) equals the expected difference between the count for (1,2) and (2,1); $(1/4)\mu_{11} - \mu_{02}$ is the

same as (1/4)$\mu_{11} - \mu_{20}$. This constraint reduces the number of parental "mating type" parameters from 9 to 6.

A third potential assumption is parental allelic exchangeability, which is when for a set of four alleles carried by a pair of parents, the alleles are randomly allocated to the two individuals; $\mu_{11} = 4\mu_{02} = 4\mu_{20}$. The exchangeability assumption implies the first two assumptions, and thus, the expected difference between the count for (1,0) and (0,1) and the expected difference between the count for (1,2) and (2,1) are equal to each other which are equal to zero. This implies a third constraint.

To impose these constraints, log-linear Poisson regression was employed as outlined by Shi et al.[213] This model takes the following form:

$$ln[E(N_{mcd})] = \theta_{mc} + \delta d + \alpha_1 d I_{m=1} + \alpha_2 d I_{m=2} + \beta_1 d I_{c=1} + \beta_2 d I_{c=2}$$

where $E(N_{mcd})$ is the expected value of the counts of families with each of maternal genotypes, child genotypes, and case or control status and d is case or control status where $d = 1$ for a case and $d = 0$ for a control; and $I_{(m=j)}$ and $I_{(c=i)}$ are indicators for whether a mother or child has j (= one or two) copies of the variant allele. The θ_{mc} parameters allow flexibility of the control-mother distribution (avoiding the need to assume Hardy-Weinberg for the source population) and ensure that the parental genotype distribution is flexible as it is only constrained by the family-based constraints. Poisson regression provides a way to impose constraints on these parameters. Here, the model is based on the natural log of the expected cell count, and subject to non-linear constraints, as described above. Those constraints can be imposed by observing that the complete data (if the fathers had also been included) would follow a log-linear model and consequently statistical missing data methods (the EM algorithm) can be used to maximize the likelihood.[235] In a triad analysis, one would use the 15 cells, but here, fathers' genotypes are missing by design. The program uses the expectation-maximization algorithm to maximize the fatherless likelihood, and enables one to impose the constraints of family structure. Missingness must be noninformative, but since all fathers are missing, one can easily make this assumption.

Using log-linear Poisson regression with the EM algorithm to impose one or several constraints of family structure should improve power compared to the logistic model. Doing so also provides an opportunity to investigate imprinting or parent of origin effects.

LEM software[236] was used to fit these models. The expectation maximization algorithm was used to incorporate dyads with missing genotypes. Likelihood ratio tests comparing reduced models with maternal genotype or child genotype with the saturated model were performed to determine p-values for both maternal and child genetic effects, each adjusted for the other genotype. A 4 degree-of-freedom likelihood ratio test was used to determine joint p-values for simultaneous tests of maternal/child genetic effects. Point estimates and 95% confidence intervals for relative risks were calculated for each SNP for both maternal and child genotype.

Confounders were identified *a priori* based on existing literature. A directed acyclic graph (DAG) was used to identify a minimally sufficient adjustment set. DAGs for maternal genotype as the exposure and fetal genotype as the exposure are shown in Figure 7. There are few factors that directly affect our exposure (SNP), therefore, there are very few that are true confounders and necessary for adjustment in the models.

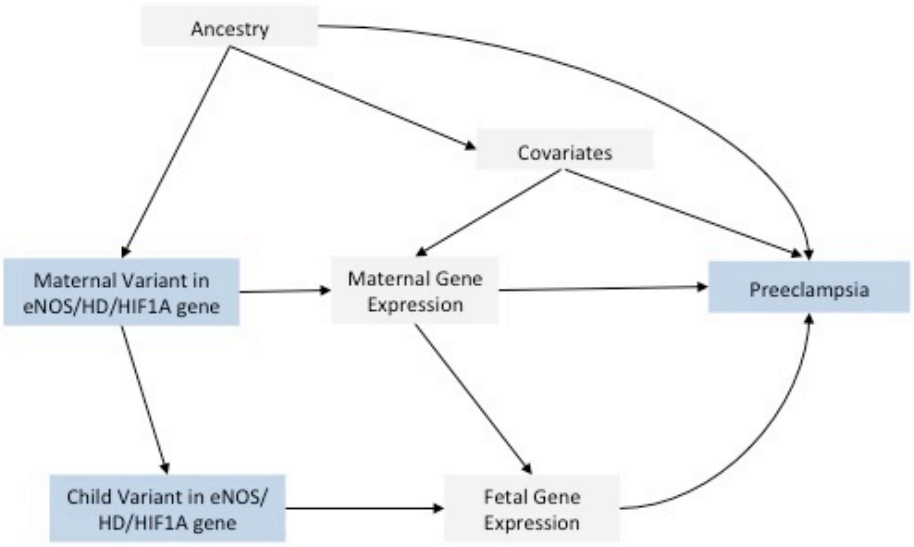


Figure 12. Directed Acyclic Graph for Aim 1

Based on the DAG, aside from correlated familial genotype, ancestry is the only other potential confounder for the main effects of genes on preeclampsia. Population stratification occurs when there are differences in allele frequencies in cases and controls due to differences in ancestry rather than in true disease risk.[156] Since all participants were required to speak Norwegian to participate in the study and Norway's population is relatively homogeneous, population stratification bias may not be a considerable concern. However, studies of other Northern European populations have indicated significant heterogeneity even among small geographic areas.[237,238] In genetic studies that employ logistic regression, eigenstrata are generally included as covariates to reduce population stratification bias, however, the use of log-linear Poisson regression with a mating type parameter in the model should reduce the need for adjustment with eigenstrata. Nevertheless, because of concern that missingness may depend on membership in subpopulations, samples that fell outside of the European ancestral population were excluded in sensitivity analyses, as described in paper 1 methods.

To visually inspect results across the whole genome, we generated quantile-quantile plots of the observed versus expected -$\log_{10}$(p-values) for each SNP from the dyad models. We also calculated the genomic inflation factor, lambda[226], for maternal and child genetic effects, which is based on comparing the median chi-squared value with its expectation under the null. Separate plots were generated for maternal and child effects.

For Aim 1b, we examined these associations by preeclampsia subtype. We performed separate analyses for early-onset preeclampsia, severe preeclampsia, preeclampsia with early delivery, and preeclampsia accompanied by small for gestational age, all defined dichotomously. The criteria and sample for each subtype is described in Table 6.

Table 6. Criteria and number of cases for preeclampsia subtypes

Phenotype	**Criteria**	**Case Pregnancies**
Preeclampsia[a]	• New onset systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure of ≥ 90 mm Hg AND • Proteinuria ≥ 0.3 g/24-hr or ≥1+ on urine dipstick	1545
Severe preeclampsia[a]	General requirements of preeclampsia plus • Systolic blood pressure of ≥ 160 mm Hg or diastolic blood pressure of ≥ 110 mm Hg OR • Proteinuria ≥ 5g/24-hr or ≥3+ on urine dipstick	308
Early-onset preeclampsia	General requirements of preeclampsia plus • Diagnosis prior to 34 completed weeks of gestation	277
Preeclampsia with early delivery	General requirements of preeclampsia plus • Delivery prior to 34 completed weeks of gestation	132
Preeclampsia with small-for-gestational-age	General requirements of preeclampsia plus • Infant born <10th percentile weight for gestational age[b]	349

[a] Cases with a validated diagnosis of eclampsia in the Medical Birth Registry of Norway were included in the preeclampsia and severe preeclampsia phenotypes.
[b] Population percentiles derived from Norwegian distribution, eSnurra Norway

Specific Aim 2 Statistical Analysis

For aim 2, we extended the log-linear Poisson regression models used in aim 1 to include interactions between genotype and smoking. Smoking was reported at several time points during pregnancy in the early and late pregnancy questionnaires and descriptive statistics were examined at each time point. Missing smoking values were analyzed for associations with the exposure and outcome to inform the likelihood of bias to missingness. Bivariate associations of smoking with each SNP and smoking with preeclampsia were analyzed. The primary goal of aim 2 was to determine if maternal smoking modified the relationship between SNP and preeclampsia. In aim 2, we extended the model in aim 1 to include two maternal and two child genetic risk parameters to saturate for codominant genetic main effects, and an interaction term was included to assess genotype by environment interactions as follows:

$$\ln[E(N_{mcde})] = \theta_{mc} + \delta d + \gamma I_e + \sigma d I_e + \alpha_1 d I_{m=1} + \alpha_2 d I_{m=2} + \beta_1 d I_{c=1} + \beta_2 d I_{c=2} + \omega d I_e \times G$$

Where $E(N_{mcde})$ is the expected value of the counts of families with each of maternal genotypes, child genotypes, case or control status, and smoking status; d is case or control status where $d = 1$ for a case and $d = 0$ for a control; $I_{(e=1)}$ is an indicator for maternal smoking; and G is the number of copies of the variant allele that the mother carries (when assessing maternal interaction) or that the child carries (when assessing child interaction). The θ_{mc} parameters allow flexibility of the control-mother distribution (avoiding the need to assume Hardy-Weinberg for the source population) and ensure that the parental genotype distribution is only constrained by the family relationships.

DAG analysis was used to determine inclusions of any covariates (Figure 13).

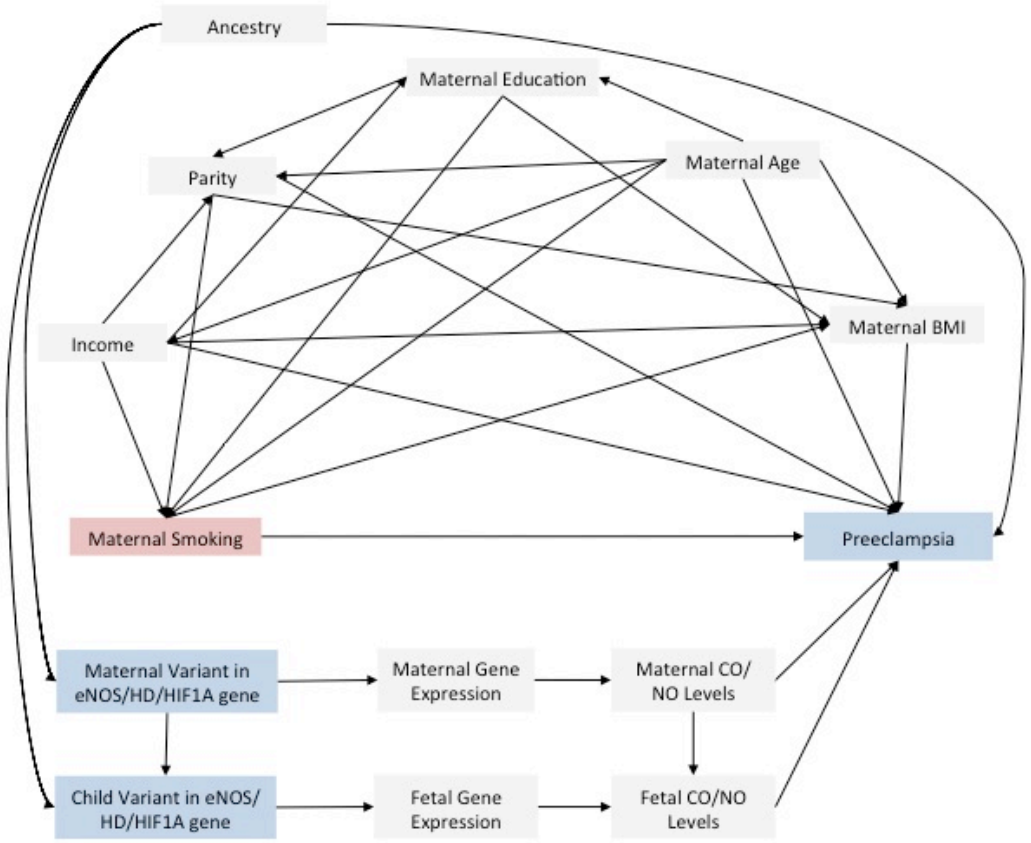


Figure 13. Directed Acyclic Graph for Aim 2

DAGs are not well-equipped to fully demonstrate effect measure modification, however, one can show covariates that may independently confound the relationship between the environmental factor and outcome (smoking and preeclampsia) and the relationship between the genetic factor and outcome. Maternal age, maternal education, maternal BMI, income, and parity may all confound the relationship between maternal smoking and preeclampsia. Ancestry may confound both the relationship between SNP and preeclampsia and smoking and preeclampsia. There is a potential concern that a gene could affect a woman's propensity to smoke, thus making smoking a mediator in the gene-outcome relationship, so this should be considered.

LEM software[236] was used to fit these models. The expectation maximization algorithm was used to incorporate dyads with missing genotypes. Likelihood ratio tests comparing reduced models without the interaction term with the saturated model were performed to determine p-values for interaction for both maternal-genotype interaction and child-genotype interaction, each adjusted for the other genotype.

Point estimates and 95% confidence intervals for relative risks were calculated for both maternal and child genotype-smoking interaction for each of the following:

1) The risk of preeclampsia for maternal smoking compared to non-smokers stratified by maternal or child genotype

2) The risk of preeclampsia for those with 1 copy or 2 copies of the variant allele compared to 0 copies stratified by maternal smoking

3) The risk of preeclampsia of maternal smoking and having 1 or 2 copies of variant allele compared to a common referent group of non-smokers with 0 copies of the variant allele.

Multiple Comparisons

Because of the large number of SNPs being examined in this study, multiple hypothesis testing is a concern because some associations would be expected due to chance alone. This number increases as the number of hypotheses tested increases.

The Bonferroni correction is frequently used to correct for multiple comparisons. The correction divides alpha by the total number of statistical tests to obtain a corrected “family-wise” Type I error rate. The Bonferroni correction, however, may be overly conservative, as it does not take into account the potential of correlation between tests. Therefore, to account for multiple comparisons, we calculated the false discovery rate (FDR), which is the expected proportion of type 1 errors (false positives) among all positive tests.[239] We used an FDR of <0.05 (reported as *Q*-values) as our threshold for considering a finding noteworthy.

Interaction tests generally have low power to reject homogeneity,[240] thus, in aim 2 we employed a more generous cutpoint that also acknowledges issues of multiple testing ($P < 0.001$). We calculated both Bonferroni corrected p-values and false discovery rate[239] but they may be overly conservative for this hypothesis-driven, yet exploratory analysis.

Sensitivity Analyses

We conducted several sensitivity analyses. As discussed, the top 3 principal components were plotted with 1000 Genomes reference populations. Cases and controls were plotted along ancestral axes to examine differential case and control status by ancestry. A sensitivity analysis was performed excluding outliers for the top principal component >0.01 and >0.04.

Sensitivity analysis were also conducted among nulliparous women only, and among overweight/obese women only.

Replication Methods

Top hits for each phenotype and one additional SNP in LD for each top hit were sent to the InterPregGen consortium for replication analysis.[241] InterPregGen samples from mother-child pairs from the United Kingdom were used for replication analysis. Cases came from the UK Genetics of Pre-eclampsia (GOPEC) consortium. The same standard definition of preeclampsia was used. Population controls came from the Wellcome Trust Case-Control Consortium 1. Maternal samples (1875 cases, 5088 controls) and child samples (1004 cases, 5286 controls) were analyzed separately for SNP associations with preeclampsia using logistic regression assuming an additive model. Samples from the GOPEC population included phenotype information for early preeclampsia, so were also analyzed as a subtype for

preeclampsia with severe features (505 maternal cases, 5051 maternal controls, 276 child cases, 5297 child controls).

GOPEC contains DNA samples from mother-baby pairs of preeclampsia recruited at diagnosis between 1992 and 2009 for genetic studies of preeclampsia in the United Kingdom.[241,242] Data for controls in GOPEC come from the WTCCC2 genome-wide analysis of UK 1958 Birth Cohort and UK National Blood Services.

Case samples were assayed on the Illumina OmniExpress chip. Standard quality control procedures were conducted with PLINK software (http://pngu.mgh.harvard.edu/~purcell/plink/) and SMARTPCA (EIGENSOFT). Samples were excluded if individual call-rate was < 95%, heterozygosity was >3 standard deviations from the mean, any of the first 3 HapMap (based on CEU, YRI, CHB, JPT and GIH) principal axes of variation were >4 standard deviations from the mean, and recorded and DNA sex were discrepant. Related individuals (IBD>0.1) with the lowest call-rate were preferentially removed. Variants were excluded if call-rate was <95%, they had deviations from Hardy-Weinberg equilibrium $P<1 \times 10^{-6}$, minor allele frequency (MAF) was <1%, and non-random missingness of uncalled genotypes was Bonferroni corrected $P < 0.05$. We used WTCCC1 population controls from the National Blood Donors Cohort and UK 1958 Birth Cohort. These samples were genotyped on the Illumina 1.2M chip and the standard QC described above was then applied. SNPs that were not genotyped in the control dataset were imputed using IMPUTE2 (impute_v2.3.0) and SHAPEIT2 using the pre-phasing workflow against the 1000 Genomes Phase 1 reference panel.

Power

Power calculations were conducted by assuming Hardy Weinberg Equilibrium to find expected cell counts for controls. The additive log-linear model was then used to estimate expected cell counts for the cases. The chi-square noncentrality statistic was compared for the null model versus the saturated model to find the study power. A sample size of 1064 case and

984 control dyads was used. The case-mother/control-mother model contains four relative risk parameters, as described in the methods. Tables 7 and 8 show power for two scenarios. For reference, a Bonferroni correction for 1,500 tests would be 3.3×10^{-5}.

Table 7. Study power for the association of genetic variants and preeclampsia (n=1064 cases, 984 controls) with relative risk parameters R_1=1.1, R_2=1.2, S_1=1.25, S_2=1.5 at varying alpha levels and minor allele frequencies.

	Type I Error			
MAF	**10^{-2}**	**10^{-3}**	**10^{-4}**	**10^{-5}**
0.10	0.55	0.28	0.12	0.05
0.20	0.82	0.59	0.36	0.19
0.30	0.90	0.72	0.50	0.30
0.40	0.92	0.75	0.54	0.34

Table 8. Study power for the association of genetic variants and preeclampsia (n=1064 cases, 984 controls) with relative risk parameters R_1=1.25, R_2=1.5, S_1=1.5, S_2=2.0 at varying alpha levels and minor allele frequencies.

	Type I Error			
MAF	**10^{-2}**	**10^{-3}**	**10^{-4}**	**10^{-5}**
0.10	>0.99	0.99	0.96	0.90
0.20	>0.99	>0.99	>0.99	>0.99
0.30	>0.99	>0.99	>0.99	>0.99
0.40	>0.99	>0.99	>0.99	>0.99

For power calculations of gene by smoking interactions, we assumed an exposure prevalence of 10%, based on a recent analysis of smoking by trimester in the MoBa cohort.[75] Smoking prevalence among MoBa mothers was 23% in the first trimester and 9% in the third trimester. Table 9 presents power calculations for a gene by smoking scenario in which I assumed relative risk parameters the same as above in the unexposed, and a doubling of effect in exposed mothers.

Table 9. Study power for the interaction of genetic variants and smoking and preeclampsia (n=1064 cases, 984 controls) with relative risk parameters R_1=1.1, R_2=1.2, S_1=1.25, S_2=1.5, RE_1=1.1, RE_2=1.2, SE_1=1.5, SE_2=2.0, and an environmental effect of 0.8 at varying alpha levels and minor allele frequencies.

	Type I Error			
MAF	**10^{-2}**	**10^{-3}**	**10^{-4}**	**10^{-5}**
0.10	0.50	0.24	0.10	0.04
0.20	0.79	0.55	0.32	0.17
0.30	0.88	0.69	0.46	0.28
0.40	0.90	0.72	0.51	0.32

As expected, power to detect interactions is much lower, and analysis of all SNPs are considered exploratory. Since this is the first study to investigate gene by smoking interactions with preeclampsia, the exploratory investigation is warranted, despite low power. A much less conservative scenario (R_1=1.25, R_2=1.5, S_1=1.5, S_2=2.0, RE_1=1.25, RE_2=1.5, SE_1=3.0, SE_2=4.0, and environmental effect=0.8) has power of >90% at each minor allele frequency and demonstrates there is reasonable power to larger effects, which is plausible given that we are predicating on a strong environmental effect.

CHAPTER IV. A FAMILY-BASED STUDY OF CARBON MONOXIDE AND NITRIC OXIDE SIGNALING GENES AND PREECLAMPSIA

Summary Paper 1

Preeclampsia is thought to originate during placentation, with incomplete remodeling and perfusion of the spiral arteries leading to reduced placental vascular capacity. Nitric oxide (NO) and carbon monoxide (CO) are powerful vasodilators that play a role in the placental vascular system. Although family clustering of preeclampsia has been observed, the existing genetic literature is limited by a failure to consider both mother and child.

We conducted a nested case-control study within the Norwegian Mother and Child Birth Cohort of 1,545 case-pairs and 995 control-pairs from 2,540 validated dyads (2,011 complete pairs, 529 missing mother or child genotype). We selected 1,518 single nucleotide polymorphisms (SNPs) with minor allele frequency >5% in NO and CO signaling pathways. We used log-linear Poisson regression models and likelihood ratio tests to assess maternal and child effects.

One SNP met criteria for a false discovery rate *Q*-value <0.05. The child variant, rs12547243 in adenylate cyclase 8 (*ADCY8*), was associated with an increased risk (RR=1.42 [95% CI: 1.20, 1.69] for *AG* vs *GG*, RR=2.14 [1.47, 3.11] for *AA* vs *GG*, *Q*=0.03). The maternal variant, rs30593 in *PDE1C* was associated with a decreased risk for the subtype of preeclampsia accompanied by early delivery (RR=0.45 [0.27, 0.75] *TC* vs *CC*; *Q*=0.02). Some evidence for association was seen for preeclampsia accompanied by small for gestational age for the child SNP, rs30562, of *PDE1C* (*Q*=0.06).

This study uses a novel approach to disentangle maternal and child genotypic effects of NO and CO signaling genes on preeclampsia.

Introduction

Preeclampsia is a common pregnancy complication, affecting approximately 2-7% of pregnant women, and typically characterized by new-onset gestational hypertension and proteinuria after 20 weeks' gestation.[2] The only definitive treatment is delivery, and it is a large contributor to medically-indicated preterm birth.[4] Preeclampsia is associated with serious maternal and fetal morbidity and mortality.[2] In the mother, preeclampsia may progress to eclampsia, and is also associated with placental abruption, thrombocytopenia, hepatic dysfunction including subcapsular liver hematoma, renal insufficiency, pulmonary edema, cerebrovascular accident, and death.[2,27,32–34] Fetuses are at risk for intrauterine growth restriction, oligohydramnios, and intrauterine fetal demise.[243] Preterm neonates incur additional risks associated with immaturity at birth and are at risk for perinatal death.[2,42,43] Currently, low-dose aspirin is the only therapy shown to reduce the risk of preeclampsia in high-risk individuals, but unfortunately this is only modestly effective (RR=0.87; 95% CI [0.79, 0.96]).[25,244]

Though incompletely understood, preeclampsia is hypothesized to originate during placentation.[2,4] Thus, both maternal and fetal components may contribute to the condition. During normal placentation, the fetal cytotrophoblast invades the maternal decidua and penetrates the walls of adjacent maternal spiral arteries. The arteries lose smooth muscle in order to increase vascular dilation. It is hypothesized that some individuals have incomplete remodeling and perfusion of the spiral arteries, leading to reduced vascular capacity[5] and placental ischemia and hypoxia,[55,57,245] leading to maternal endothelial dysfunction and the subsequent clinical symptoms of preeclampsia.[245]

Nitric oxide (NO) and carbon monoxide (CO) are powerful vasodilators[7,8] that may play important roles in the etiology of preeclampsia by increasing vascular capacity during spiral

artery remodeling. NO and CO are produced by combustion such as in cigarette smoke or vehicle exhaust,[9,10] but are also produced endogenously in the body.[11–16] The well-known inverse relationship of maternal smoking with preeclampsia[71] may in part be attributable to the vasodilatory effects of NO and CO. There is evidence that combustion products may be required, as a large Swedish study found *increased* risk in users of smokeless tobacco.[246] NO and CO are associated with smooth muscle relaxation and blood pressure regulation[12,14] and there is evidence of their role specifically in the placental vascular system.[15,77]

Genomic studies can help identify etiologically relevant underlying biologic pathways. Although the reason for the inverse association between maternal smoking and preeclampsia is unknown, assessing variants in genes that influence the endogenous production of CO and NO, which are components of cigarette smoke, may provide evidence for a potential mechanism of action. There is evidence for genetic susceptibility for preeclampsia. Numerous studies have suggested a familial predisposition for preeclampsia risk in first-degree relatives of women with preeclampsia[123] and studies of familial aggregation of preeclampsia have estimated heritability to be as much as 50%.[124–126] Although these estimates have been supported by both large-scale registry-based epidemiologic studies[127,128] and studies of transgenic mice,[123,129] findings from human genetic studies have been inconsistent, and few genetic associations have been found. Both maternal and child genetics may contribute to risk of preeclampsia.

Our objective was to determine if maternal or child single nucleotide polymorphisms (SNPs) in NO and CO signaling pathways were associated with preeclampsia using a mother-child dyad design, nested within the Norwegian Mother and Child cohort (MoBa). We examined SNPs within three canonical pathways important for both CO and NO activity. Exploring both maternal and child genotype and identifying variants that may play a role in both endogenous and exogenous NO and CO may help establish potential therapeutic targets for this serious and life-threatening condition.

Methods

Study population

This study is a nested case-control study within the Norwegian Mother and Child Cohort Study (MoBa), conducted by the Norwegian Institute of Public Health.[222] MoBa is a large prospective birth cohort of pregnant women and their offspring, recruited throughout Norway from 1999 to 2008 (N=112,908 pregnancies). All pregnant women living in Norway who gave birth at a hospital or maternity unit with more than 100 births annually and who could speak Norwegian were eligible; MoBa investigators applied no other exclusion criteria. Pregnant women were recruited by mail prior to their routine ultrasound appointment at 17 to 20 weeks' of gestation. Of all women invited to participate, 41% enrolled in the study.[222] Participants completed two prenatal questionnaires about their health and environment. Survey completion rate was 91% for the early pregnancy questionnaire (administered between weeks 13 and 17) and 83% for the late pregnancy questionnaire (administered at week 30).[222] Maternal blood was collected at the first ultrasound appointment and cord blood was collected at birth. Maternal blood was received from 89% of participants and child (cord) blood from 81% of children in the cohort.[222] DNA was extracted at the time of collection and then stored at the MoBa Biobank.

Birth outcome information was obtained from the Medical Birth Registry of Norway.[247] For the purposes of this analysis, we included women with a singleton pregnancy who conceived spontaneously, were verified cases or controls, returned both the early and late pregnancy questionnaires, and had no history of chronic hypertension. Preeclampsia case/control status was verified using antenatal records and hospital discharge codes, as previously described.[223] Preeclampsia was defined using American College of Obstetrics and Gynecologists (ACOG) criteria (see below for additional details).[27] All observations registered as preeclampsia cases and a random sample of 2000 pregnancies registered as being unaffected by preeclampsia were selected from MoBa to be verified. Of the 3500 registered preeclampsia cases and 1840 registered to be unaffected by preeclampsia for which records were received,

2936 pregnancies identified as preeclampsia cases from the MBRN were verified to have been affected by preeclampsia, and 1745 pregnancies identified as unaffected by preeclampsia were found to be negative for preeclampsia. In total, 2682 preeclampsia case samples (2236 samples from 1118 mother/child pairs and 446 unpaired maternal samples) and 1967 non-preeclampsia control samples (1936 samples from 968 mother/child pairs and 31 unpaired maternal samples) met inclusion criteria and were genotyped. There were no unpaired offspring. (Figure 7).

Outcome Assessment

Details of the validation study for preeclampsia have been previously described[223], and information relevant to the current investigation is included in Supporting Information Methods. In brief, "preeclampsia" was as defined by the American College of Obstetrics and Gynecologists (ACOG),[27] which specifies that both of the following be present:

4) Systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure of ≥ 90 mm Hg occurring after 20 weeks' gestation in a woman whose blood pressure has been previously normal, and
5) Proteinuria, with excretion of ≥ 0.3 g of protein in a 24-hour urine specimen or as measured by 1+ on urine dipstick.

A 2013 revision of this definition also included clinical symptoms, however, we used the criteria in use at the time of validation, i.e. the above two conditions. We also considered as "true cases" any case with an ICD-10 code of severe PE (O14.1) or HELLP syndrome (O14.2) with delivery < 37 weeks, or ICD-10 code of eclampsia (O15), which are routinely validated by the MBRN. Criteria for preeclampsia subtypes are presented in Table 6.

Gene and SNP Selection

For this study, 66 genes (Table S1) involved in CO and NO activity were identified from three canonical pathways, which included: 1) endothelial nitric oxide synthase (eNOS) signaling

pathway, which accomplishes the synthesis of NO from L-arginine, 2) heme degradation, which accomplishes the breakdown of hemoglobin into CO and bilirubin and 3) hypoxia-inducible factor 1-alpha (HIF1A), which regulates oxygen homeostasis and response to hypoxia. A total of 1,518 SNPs were selected and analyzed for this study using a 10kb upstream and downstream margin around the transcription start and end sites for each gene.

DNA Genotyping and Quality Control

SNPs were genotyped by the UNC Mammalian Genotyping Core using the HumanCoreExome+ array from Illumina (Illumina, Inc., San Diego, CA). Samples and SNPs were examined using PLINK 1.07 (http://pngu.mgh.harvard.edu/purcell/plink) for quality control. SNPs were excluded if the missing rate exceeded 5% or there was substantial deviation from Hardy-Weinberg Equilibrium ($p < 1x10^{-3}$) using PLINK. We also excluded SNPs with a minor allele frequency <5%. Samples with known genotype (n = 84) and DNA replicates (n = 51) were included on each plate and generally exhibited high accuracy; however, one pair of discordant duplicate samples was excluded. All other quality control samples on the implicated plates were found to be perfectly concordant. All subject-specific call rates were acceptable (minimum 97.2%). Sex-specific markers were inspected and 3 samples with sex discrepancies were excluded. Parent-child relatedness and inbreeding within the cohort was confirmed by identity by descent. Thirteen mother-child pairs were dropped because relatedness could not be confirmed (expected pi-hat = 0.5, observed pi-hat<0.128). Additionally, 16 pairs of siblings or cousins among the mothers were flagged as related (pi-hat>0.125). For each such pair of related mothers, we preferentially included the parent-child pair with the most complete genetic data, or in the case of equivalence, randomly sampled between them.

Quantile-quantile plots and calculation of genomic control lambda[226] indicated no systematic test statistic inflation, suggesting that population stratification was negligible (Figure 9). However, the top 3 principal components of genetic variation were plotted for the MoBa data

together with the 1000 Genomes reference populations and visually inspected to assess evidence of admixture (Figure 10). A sensitivity analysis excluding varying levels of population outliers was performed and did not substantially influence results. (Table S11). The final analysis sample (n=4551 total samples) included dyads with both mother and child genotype data as well as incomplete dyads with only mother or child genotype data (n=2621 preeclampsia case samples [1076 mother/child pairs, 459 mother only, 10 child only], n=1930 control samples [935 mother/child pairs, 46 mother only, 14 child only]). Supporting Information Figure 7 describes the sample selection and quality control process.

Statistical Analysis

To simultaneously account for maternal and child genotype, we used the case-mother control-mother log-linear modeling approach proposed by Shi et al.[213] This method uses Poisson regression to model expected counts of each possible genetic mating type combination under the assumption of Mendelian inheritance. This method allows one to account for the correlation between maternal and child genotype and improves power compared to a logistic model.[212,213,248] Two maternal and two child genetic risk parameters were included in the model to saturate for codominant genetic main effects, as follows:

$$\ln[E(N_{mcd})] = \theta_{mc} + \delta d + \alpha_1 dI_{m=1} + \alpha_2 dI_{m=2} + \beta_1 dI_{c=1} + \beta_2 dI_{c=2}$$

Where $E(N_{mcd})$ is the expected value of the counts of families with each of maternal genotypes, child genotypes, and case or control status; d is case or control status where $d = 1$ for a case and $d = 0$ for a control; and $I_{(m=j)}$ and $I_{(c=i)}$ are indicators for whether a mother or child has j (= one or two) copies of the variant allele. The θ_{mc} parameters allow flexibility of the control-mother distribution (avoiding the need to assume Hardy-Weinberg for the source population) and ensure that the parental genotype distribution is only constrained by the family relationships.

LEM software[236] was used to fit these models. The expectation maximization algorithm was used to incorporate dyads with missing genotypes. Likelihood ratio tests comparing

reduced models with maternal genotype or child genotype with the saturated model were performed to determine p-values for both maternal and child genetic effects, each adjusted for the other genotype. A 4 degree-of-freedom likelihood ratio test was used to determine joint p-values for simultaneous tests of maternal/child genetic effects. Point estimates and 95% confidence intervals for relative risks were calculated for each SNP for both maternal and child genotype.

In genetic studies that employ logistic regression, eigenstrata are generally included as covariates to reduce population stratification bias, however, the use of log-linear Poisson regression with a mating type parameter in the model should reduce the need for adjustment with eigenstrata. Nevertheless, because of concern that missingness may depend on membership in subpopulations, samples that fell outside of the European ancestral population were excluded in sensitivity analyses, as described above.

To account for multiple comparisons, we calculated the false discovery rate (FDR), which is the expected proportion of type 1 errors (false positives) among all positive tests.[239] We used an FDR of <0.05 (reported as *Q*-values) as our threshold for considering a finding noteworthy.

To visually inspect results across the whole genome, we generated quantile-quantile plots of the observed versus expected $-\log_{10}$(p-values) for each SNP from the dyad models (Figure 14). We also calculated the genomic inflation factor, lambda[226], for maternal and child genetic effects, which is based on comparing the median chi-squared value with its expectation under the null. Separate plots were generated for maternal and child effects.

Replication Methods

All SNPs with *Q*-values < 0.2 for both preeclampsia overall and preeclampsia sub-phenotypes were sent to the InterPregGen consortium for attempted replication analysis.[241] Within InterPregGen, cases came from the UK Genetics of Pre-eclampsia (GOPEC)

consortium. The same standard definition of preeclampsia defined cases. Population controls came from the Wellcome Trust Case-Control Consortium 1. Maternal samples (1875 cases, 5088 controls) and child samples (1004 cases, 5286 controls) were analyzed separately for SNP associations with preeclampsia using logistic regression, assuming a logit-additive model. A subset of preeclampsia cases included phenotype information for early preeclampsia, so were also analyzed as a proxy for the subtypes of preeclampsia with additional complications (505 maternal cases, 5051 maternal controls, 276 child cases, 5297 child controls). Complete methods for recruitment, genotyping, and quality control of the replication sample are described Chapter III.

Results

The final analysis sample consisted of 4,551 individual samples for 2,011 complete mother-child dyads, 505 samples with only maternal genotype data (459 cases), and 24 children with only child genotype data (10 cases) (n=2,540 pregnancies) (Table 10). Mean maternal age was 29.6 years (SD 4.7) and most had a university degree (55.3%). As expected, a greater proportion of women with preeclamptic pregnancies were nulliparous and of high body mass index (overweight or obese) compared to those without preeclampsia (66% and 46% compared to 41% and 28%, respectively). Babies born to women with preeclampsia were more often preterm (21.5% compared to 3.3%) and small-for-gestational-age (SGA) (22.6% compared to 7.7%) than controls. Severe preeclampsia (including eclampsia) was present in 20% of women with preeclampsia.

Results of tests for maternal genotypic associations controlling for child genotype are summarized in Figure 14a and for child genotypic associations controlling for maternal genotype are summarized in Figure 14b. In the joint 4-degree of freedom test, we found one SNP to be significant ($Q \leq 0.05$), however, this SNP was only individually significant in the child and not in

the mother. We found a child association of increasing risk in the variant allele of this SNP (rs12547243, MAF = 0.29), a synonymous substitution in a coding region of *ADCY8* on chromosome 8. The estimated relative risk (RR) and 95% confidence interval (CI) was 1.42 [1.20, 1.69] for 1 copy of the minor allele and 2.14 [1.47, 3.11] for 2 copies, *Q*=0.04) (Table 11). Although no maternal genotypic associations met our FDR threshold, there were a number of suggestive maternal genotypic associations for SNPs in *ESR1, PDE1C, PIK3C2G*, and *GUCY1A3*. Generally, the *ESR1* and *PDE1C* SNPs were associated with a reduced risk of preeclampsia and the *PIK3C2G* and *GUCY1A3* SNPs were associated with an increased risk of preeclampsia, however, few showed a dose-response pattern and risk ratios were mostly null for the homozygous genotype. Table 11 shows both mother and child effect estimates, and the joint test results, for all top SNPs with FDR *Q*-values $\leq$0.20.

Results of a sensitivity analysis excluding population outliers along axes of ancestral variation are presented for all SNPs with FDR $Q \leq 0.20$, although only 5 are significant at 0.05. Supporting Information Table S11a provides risk ratios and test results for exclusion of 17 pairs of observations with first principal component >0.04 (See Figure 12) and Table S11b provides these estimates for exclusion of 47 pairs of observations with first principal component >0.01. Exclusion of these observations did not substantially influence results; effect estimates were similar and the p-values were stable.

Because preeclampsia is a heterogeneous condition for which underlying etiologies may differ, we also repeated the analysis within preeclampsia subtypes. Results for all associations for each subtype with $Q \leq 0.2$ are presented in Table 4. Within subtypes, we found associations for two SNPs within *PDE1C.* We found a maternal association of rs30593 (MAF=0.35) for preeclampsia accompanied by delivery before 34 weeks' gestation (RR=0.45 [0.27, 0.75] for 1 copy; RR=1.44 [0.63, 3.30] for 2 copies; *Q*=0.02). We also found a suggestive child association

of rs30562 (MAF=0.35) for preeclampsia accompanied by SGA (RR=0.50 [0.35, 0.71] for 1 copy; RR=1.00 [0.60, 1.67] for 2 copies; *Q*=0.06).

We provided the SNPs with *Q*-values < 0.2 overall and within subtypes to the InterPregGen Consortium for analysis. Because we found different lead SNPs among the general preeclampsia phenotype and preeclampsia subtypes, in the replication sample we assessed these SNPs for both associations with general preeclampsia (Table S12a) and early preeclampsia (Table S12b), the only sub-phenotype for which we had replication data. Our lead SNP for preeclampsia with early delivery, a maternal association of rs30593 in *PDE1C,* was nominally associated with preeclampsia, but in the UK GWAS child population (uncorrected p=0.05). As with this SNP in the MoBa study, we saw a similar reduced risk of preeclampsia (RR=0.90; 95% CI [0.82, 1.00]) in the replication study. Complete replication results are reported in Table S12.

Discussion

In this large, case-control study of maternal and child genetic variants in CO and NO pathways and validated preeclampsia, we report potential genetic associations of interest within both the mother and the child. In the child, we found a novel association of rs12547243 in *ADCY8*, a SNP with no previously reported associations. We found increasing risk of preeclampsia with the *AG* and *AA* genotypes compared with the *GG* genotype. Although we found no maternal genotypic associations that met our FDR threshold, we did find some suggestive maternal genotypic associations (*Q* < 0.1) for SNPs in *ESR1, PDE1C, PIK3C2G*, and *GUCY1A3*. Among preeclampsia subtypes, we found a decreased risk associated with the TC maternal genotype compared to CC genotype of rs30593 in PDE1C for preeclampsia accompanied by early delivery and a suggestive association of reduced risk associated with the child *TC* genotype compared to *CC* for rs30562 in *PDE1C* for preeclampsia accompanied by small for gestational age.

Of the lead SNPs that were sent for replication analysis, only rs30593 in *PDE1C* was nominally associated with preeclampsia in the replication sample (p=0.05), however, none of the replication SNPs were significant after adjustment for multiple comparisons. It is plausible that some of our associations failed to replicate due to both different analytic methods and a somewhat different outcome assessment in our primary analysis and replication datasets. While both populations used the same definition of preeclampsia, our MoBa analysis used only validated cases and non-cases, whereas the InterPregGen Consortium used validated cases but population controls that may or may not be pregnant women. Additionally, in our analysis, we included both mother and child genotypes and both were modeled codominantly with a genetic mating type parameter to account for mother-child family structure; by contrast, the replication analysis independently modeled mother and child genotype logit-additively. To investigate how a similar type of analysis might alter our results, we analyzed our maternal and child samples using standard logistic regression analysis modeled additively in PLINK. Our associations with rs12547243 in the child were stronger, but other associations were attenuated (data not shown).

Although not replicated, we did find some interesting associations that have been supported by animal studies and warrant further study. Several SNPs of interest are located in the genes *ADCY8, GUCY1A2,* and *PDE1C. ADCY8* encodes adenylate cyclase 8 which catalyzes the conversion of adenosine-5'-triphosphate (ATP) to 3', 5'-adenosine monophosphate (cAMP).[249] *GUCY1A2* encodes the guanylate cyclase 1 alpha 2 subunit, which acts similarly to *ADCY8* in that it catalyzes the conversion of guanosine triphosphate (GTP) to 3', 5'-adenosine monophosphate (cGMP).[250] *PDE1C* encodes calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1C and catalyzes conversion of cAMP and cGMP to their 5-prime-monophosphates (AMP and GMP).[251] All of these genes are part of the canonical pathway for cellular effects of sildenafil ("Viagra"), a phosphodiesterase inhibitor that operates through nitric oxide signaling to increase vasodilation by inhibiting conversion of cGMP to

GMP.[252] Mouse studies have demonstrated resolution of the preeclampsia phenotype with the administration of sildenafil.[253,254] and clinical trials are now underway for its investigation for treatment of preeclampsia and fetal growth restriction.[255–257]

Our study is one of the largest genetic studies of preeclampsia to date, with 1,076 case pairs (2,152 samples) and 935 control pairs (1,870 samples), which allowed us to address both mother and child genotype. In contrast to some small prior studies in ethnically diverse populations[22–24], we did not identify any associations within SNPs in *NOS2* or *NOS3*, two of the most widely-studied candidate genes for preeclampsia in the CO and NO pathways. SNPs were selected using a 10kb upstream and downstream margin around the transcription start and end sites for each gene. Although we selected these margins to ensure that we captured the entire coding region of the gene and proximate regulatory regions, recent studies of suggest that distal intergenic variants can alter gene expression through their role within the regulatory regions.[258] Candidate gene studies have frequently been inconsistent and unreplicated. This study aimed to improve upon prior candidate gene studies by including more SNPs while continuing to apply a hypothesis-driven approach to maintain statistical power.

Given the suspected pathophysiology underlying preeclampsia, it is biologically plausible that both maternal and child genotypes contribute to this pregnancy complication. Our dyad analysis accounted for family structure and included both mother and child DNA, each adjusting for the other. This method has been implemented for other child phenotypes in which both maternal and child genotype may play a role[259], but has not as well explored for pregnancy conditions. Although a few other genetic studies of preeclampsia have addressed child genotype[124,194,260], to our knowledge, none have simultaneously modeled both mother and child genotype.

While addressing limitations of previous studies by accounting for both maternal and child genetic components, this study also took on the challenge of heterogeneity of the preeclampsia phenotype. Preeclampsia was verified by antenatal medical records and hospital

diagnostic codes through medical record validation.[223] All cases of preeclampsia in our study were verified as having preeclampsia and all controls were verified as being free of preeclampsia. We were also able to classify cases into preeclampsia subtypes that may have differing underlying etiologies.

Although our analytic method allowed us to control for paired genotypes, which is likely the biggest confounder, a limitation of this study design is the inability to control for external confounders, such as by using principal components to adjust for population admixture. We addressed this limitation by assessing population stratification in the quality control process to determine if ancestral homogeneity was a plausible assumption within this Norwegian population. Comparing our population with the 1000 Genomes reference populations, we identified a handful of MoBa participants who clustered with Chinese, Amerindian, and Nigerian ancestral populations. Excluding such observations in a sensitivity analysis, however, did not substantially change our results, indicating that population stratification bias is not a serious issue. However, our results are most generalizable to populations of European descent. Additionally, despite being the largest study, we are still underpowered to look at preeclampsia subtypes. Although we do see some interesting trends and suggested associations, larger studies are needed to study these associations in greater detail.

Our results underscore the importance of addressing the contribution of both maternal and child genotype. Among our top hits, we found either maternal or child associations, but not both (e.g. a positive maternal association but null child association). Additionally, although they did not meet our significance threshold, there were instances in which we saw trends of positive associations for either the mother or child and negative associations in the other. We also found differing patterns among mothers and offspring in our subtype analyses. For example, all associations of interest for preeclampsia with early delivery were maternal, while all associations of interest for preeclampsia associated with small for gestational age were in the

child. Failure to acknowledge both maternal and child genotype may result in missing important associations in which mother and child genotype may operate differently.

In conclusion, this study uses a novel study design to disentangle maternal and child genotypic effects of NO and CO signaling genes on preeclampsia. We provide further evidence of a plausible biologic pathway and the role of NO signaling in the development of preeclampsia and support for the continuation of trials of sildenafil as a potential therapeutic treatment. Future research of genetics and preeclampsia should continue to incorporate maternal and child genetic components and expand to explore maternal-child genotypic interactions as well as interactions with exogenous sources of CO and NO.

Tables

Table 10. Demographic characteristics of pregnancies in the final study sample for Aim 1 (n=2540 pregnancies, 4551 samples)

	Total (N = 2540)		Preeclampsia Cases (N = 1545)		Controls (N = 995)	
Maternal Age (mean(SD), years)	29.6 (4.7)		29.3 (4.9)		30.1 (4.4)	
	No.	**%**	**No.**	**%**	**No.**	**%**
Maternal Education						
< High School	205	8.1	132	8.5	73	7.3
High School Graduate	703	27.7	439	28.4	264	26.5
University Degree	1405	55.3	831	53.8	574	57.7
Missing	227	8.9	143	9.3	84	8.4
Body Mass Index (kg/m^2)						
Underweight (<18.5)	57	2.2	25	1.6	32	3.2
Normal weight (18.5-24.9)	1341	52.8	712	46.1	629	63.2
Overweight (25.0-29.9)	630	24.8	444	28.7	186	18.7
Obese (30.0+)	359	14.1	264	17.1	95	9.6
Missing	153	6.0	100	6.5	53	5.3
Maternal Smoking						
Smoking in weeks 11-20	215	8.5	120	7.8	95	9.6
Missing	127	5.0	83	5.4	44	4.4
Smoking in third trimester	135	5.3	71	4.6	64	6.4
Missing	245	9.7	159	10.3	86	8.6
Nulliparous	1419	55.9	1012	65.5	407	40.9
Preterm (< 37 weeks)	365	14.4	332	21.5	33	3.3
Small for gestational age (SGA) (< 10th percentile)[a]	426	16.8	349	22.6	77	7.7
Preeclampsia subtypes						
Severe	308	12.1		19.9		
Onset <34 weeks	277	10.9		17.9		
Delivery <34 weeks	132	5.2		8.5		
Accompanied by SGA[a]	349	13.7		22.6		

[a] Population percentiles derived from Norwegian distribution, eSnurra Norway

Table 11. Summary of SNPs with FDR $Q \leq 0.2$ for tests of maternal genetic effects, adjusting for child genotype, and child genetic effects, adjusting for maternal genotype.

						Mother			Child			Joint test	
Marker[a]	**Chr**	**Position**	**MAF**	**Gene**	**Genotype**	**RR (95% CI)**	**P value**	**FDR Q value**	**RR (95% CI)**	**P value**	**FDR Q value**	**P value**	**FDR Q value**
rs7435347	4	156654735	0.17	*GUCY1A3*	*AA*	Referent	6.17×10^{-4}	0.09	Referent	0.09	0.76	3.36×10^{-3}	0.30
					GA	1.47 (1.20, 1.80)			0.83 (0.69, 0.99)				
					GG	1.00 (0.63, 1.58)			0.67 (0.40, 1.14)				
rs1569788	6	152328616	0.30	*ESR1*	*TT*	Referent	3.02×10^{-4}	0.09	Referent	0.61	0.88	2.66×10^{-3}	0.30
					CT	0.70 (0.58, 0.84)			1.07 (0.89, 1.28)				
					CC	0.93 (0.67, 1.30)			1.18 (0.83, 1.68)				
rs3020366	6	152368758	0.37	*ESR1*	*TT*	Referent	5.45×10^{-4}	0.09	Referent	0.40	0.88	4.49×10^{-3}	0.30
					CT	0.70 (0.58, 0.84)			1.11 (0.92, 1.32)				
					CC	0.86 (0.64, 1.15)			1.23 (0.90, 1.67)				
rs6462324	7	32120897	0.40	*PDE1C*	*CC*	Referent	2.05×10^{-4}	0.09	Referent	0.32	0.88	1.23×10^{-3}	0.25
					AC	0.69 (0.57, 0.84)			1.04 (0.87, 1.25)				
					AA	0.93 (0.69, 1.23)			1.24 (0.92, 1.69)				
rs6470860	8	131905190	0.41	*ADCY8*	*AA*	Referent	0.96	0.98	Referent	3.12×10^{-4}	0.15	9.88×10^{-5}	0.07
					GA	1.03 (0.85, 1.24)			1.32 (1.10, 1.58)				
					GG	1.02 (0.77, 1.35)			1.85 (1.36, 2.51)				
rs12547243	8	131921956	0.29	*ADCY8*	*GG*	Referent	0.09	0.89	Referent	2.69×10^{-5}	0.04	6.27×10^{-7}	8.78×10^{-4}
					AG	1.14 (0.94, 1.37)			1.41 (1.19, 1.67)				
					AA	0.82 (0.58, 1.17)			2.12 (1.46, 3.07)				

						Mother			Child			Joint Test	
Marker[a]	**Chr**	**Position**	**MAF**	**Gene**	**Genotype**	**RR (95% CI)**	**P value**	**FDR Q value**	**RR (95% CI)**	**P value**	**FDR Q value**	**P value**	**FDR Q value**
rs7459573	8	131928401	0.34	*ADCY8*	*AA*	Referent	0.92	0.98	Referent	1.61×10^{-4}	0.11	3.14×10^{-4}	0.11
					GA	0.97 (0.81, 1.17)			1.43 (1.20, 1.70)				
					GG	0.94 (0.69, 1.28)			1.73 (1.25, 2.38)				
rs17475920	12	18478126	0.12	*PIK3C2G*	*AA*	Referent	9.39×10^{-4}	0.10	Referent	0.13	0.80	7.12×10^{-3}	0.34
					TA	1.52 (1.21, 1.90)			0.85 (0.69, 1.04)				
					TT	1.65 (0.80, 3.40)			0.54 (0.27, 1.07)				
rs9634063	12	18593252	0.15	*PIK3C2G*	*CC*	Referent	2.70×10^{-4}	0.09	Referent	0.04	0.76	1.95×10^{-3}	0.30
					TC	1.54 (1.25, 1.91)			0.80 (0.66, 0.97)				
					TT	1.39 (0.75, 2.60)			0.56 (0.30, 1.06)				

[a] SNPs that also met $Q \leq 0.2$ and were in high linkage disequilibrium ($R^2 > 0.8$ using 1000 Genomes Pilot 1 CEU data): rs3796578 ($R^2 = 1.0$ with rs7435347); rs722208 ($R^2 = 1.0$ with rs1569788); rs3020365 ($R^2 = 0.93$ with rs3020366); rs11044095, rs1447406, rs11044129, rs7311726, rs1447408 ($R^2 \geq 0.92$ with rs9634063)

Table 12. Summary of SNPs with FDR $Q \leq 0.2$ for tests of maternal genetic effects, adjusting for child genotype, and child genetic effects, adjusting for maternal genotype for preeclampsia subtypes.

						Mother			Child			Joint test	
Marker*	**Chr**	**Position**	**MAF**	**Gene**	**Genotype**	**RR (95% CI)**	**P value**	**FDR Q value**	**RR (95% CI)**	**P value**	**FDR Q value**	**P value**	**FDR Q value**
Severe Preeclampsia													
No SNPs with $Q \leq 0.2$													
Early-onset preeclampsia (diagnosis <34 weeks)													
No SNPs with $Q \leq 0.2$													
Preeclampsia with delivery <34 weeks													
rs30593	7	32105096	0.35	*PDE1C*	*CC*	Referent	1.69×10^{-5}	0.02	Referent	0.40	0.75	5.38×10^{-5}	0.07
					TC	0.45 (0.27, 0.75)			0.63 (0.31, 1.30)				
					TT	1.44 (0.63, 3.30)			0.89 (0.27, 2.89)				
rs12785615	11	106869624	0.14	*GUCY1A2*	*GG*	Referent	6.32×10^{-4}	0.20	Referent	0.66	0.75	2.80×10^{-3}	0.62
					CG	0.63 (0.35, 1.15)			0.85 (0.35, 2.08)				
					CC	4.31 (1.42, 13.10)			--				
rs1455590	11	106869973	0.12	*GUCY1A2*	*GG*	Referent	3.50×10^{-4}	0.15	Referent	0.44	0.75	6.10×10^{-4}	0.34
					AG	0.63 (0.38, 1.03)			1.43 (0.73, 2.80)				
					AA	2.50 (1.16, 5.37)			0.65 (0.07, 5.70)				
rs11636443	15	52319696	0.43	*MAPK6*	*GG*	Referent	3.17×10^{-4}	0.15	Referent	0.21	0.75	1.89×10^{-3}	0.62
					AG	0.68 (0.42, 1.08)			0.55 (0.27, 1.11)				
					AA	1.98 (0.97, 4.02)			0.47 (0.17, 1.32)				
Preeclampsia with small for gestational age													
rs30562	7	32065264	0.35	*PDE1C*	*CC*	Referent	0.04	0.96	Referent	4.42×10^{-5}	0.06	2.76×10^{-5}	0.04
					TC	1.40 (1.04, 1.88)			0.50 (0.36, 0.71)				
					TT	1.04 (0.62, 1.73)			1.02 (0.62, 1.70)				

Figures

Figure 14a.

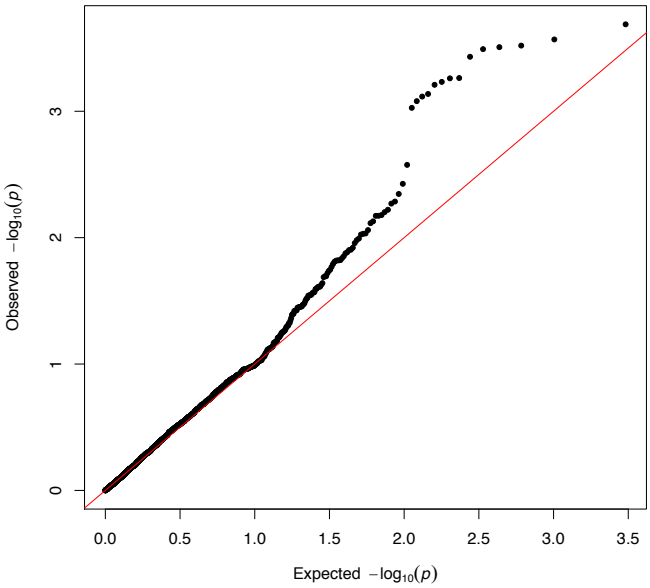


Figure 14b.

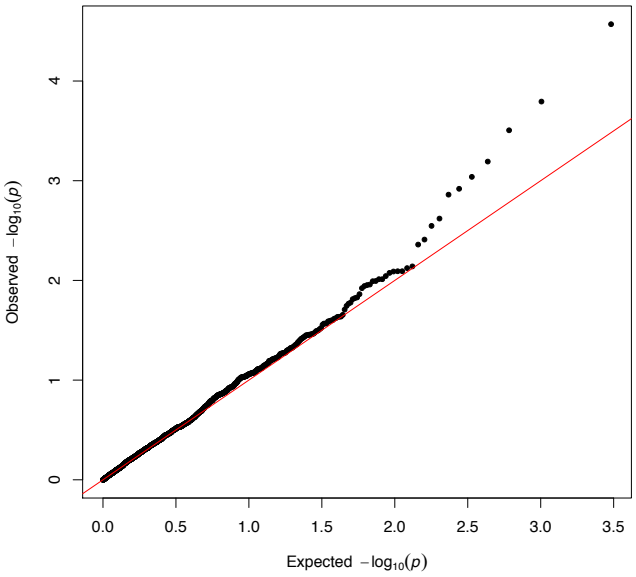


Figure 14. Quantile-quantile plots for maternal and child genotypic effects. Test results (observed -$\log_{10}$p values) are plotted against the expected -$\log_{10}$p values for each of 1,518 SNPs across 66 loci in the sample. Figure 14a. Q-Q plot of maternal genotypic effects, adjusting for child genotype. Genomic inflation factor, lambda=1.18. Figure 14b. Q-Q plot of child genotypic effects, adjusting for maternal genotype. Genomic inflation factor, lambda=1.09.

CHAPTER V. INVESTIGATION OF THE INVERSE SMOKING-PREECLAMPSIA RELATIONSHIP THROUGH THE EXPLORATION OF GENE BY SMOKING INTERACTIONS

Summary Paper 2

One of the most well-established associations with preeclampsia is the enigmatic inverse association of maternal smoking and preeclampsia. The reason for this association remains unknown. A plausible biological explanation for this relationship is through response to cigarette smoke components, either through action of vasodilators or activation of smoking detoxification pathways. Examining genes involved in these processes and their modification could provide support for a genetic or related biological mechanism.

We conducted a nested case-control study within the Norwegian Mother and Child Birth Cohort of 1,533 case-pairs and 982 control-pairs from 2,596 mother-child dyads (1,988 complete pairs, 538 missing mother or child genotype). We selected 1,915 SNPs in pathways involved in nitric oxide and carbon monoxide signaling or smoking detoxification and established smoking status from maternal questionnaire during gestational weeks 11 to 20. We used log-linear Poisson regression models with a SNP by smoking interaction term and likelihood ratio tests to assess genotype-smoking interactions.

We found limited evidence for multiplicative SNP by smoking interaction after correction for multiple comparisons. When examining effect measure modification by smoking, among our three most noteworthy SNPs (p-interaction <0.001), we saw a null association among non-smokers and a reduced risk among smokers with the variant allele for rs3765692 (*TP73*) and rs1077343 (*PIK3C2G)* and an increased risk among smokers for rs2278361 (*APAF1*).

Our findings to not provide support that the inverse smoking-preeclampsia relationship is due to a genetic effect. Dyad methods and gene-environment interaction analysis may be useful for the study of pregnancy outcomes, particularly preeclampsia.

Introduction

Preeclampsia is a common pregnancy complication, affecting approximately 2-7% of pregnant women, and is typically characterized by new-onset gestational hypertension and proteinuria after 20 weeks' gestation.[2] Preeclampsia is associated with serious maternal and fetal morbidity and mortality[2] and there are limited options for treatment other than delivery of the baby.[4]

Few risk factors for preeclampsia have been consistently identified across studies,[2] but one of the most consistent associations is the paradoxical and poorly understood inverse relationship between maternal smoking and preeclampsia. Maternal smoking is associated with as much as a 50% reduced risk of preeclampsia despite increasing risk of other poor pregnancy outcomes, such as preterm birth and fetal growth restriction which often co-occur with preeclampsia.[71–73] The reason for this association remains unknown; both biological causes[6,111] and methodological[261] reasons (e.g. survival bias) have been suggested.

One plausible biological explanation for the reduced risk of preeclampsia in women who smoke during pregnancy is the response to components of cigarettes and could reflect several mechanisms, including: 1) increased vasodilation of the spiral arteries during placentation, or 2) activation of receptors involved in detoxification that reduce endothelial dysfunction and increase angiogenesis. Both nitric oxide and carbon monoxide are components of cigarette smoke that are also produced endogenously in the body. CO and NO have been shown to be associated with smooth muscle relaxation and blood pressure regulation[12,14], and there is evidence of their role specifically in the placental vascular system.[15,77] Exposure to exogenous CO and NO may reduce risk of preeclampsia by increasing vascular dilation at a critical time

during placentation when the fetal trophoblasts are penetrating and spiral arteries are remodeling. Nicotine and polycyclic aromatic hydrocarbons are both components of cigarette smoke that activate receptors involved in smoking detoxification. Nicotine stimulates the nicotinic acetylcholine receptor and restores proangiogenic functions to endothelial cells harmed by soluble fms-like tyrosine kinase (s-flt1) and soluble endoglin (sEng) while stimulating placental growth factor (PlGF)[94] in vitro; these factors (s-flt1, sEng, PlGF) are hypothesized to be involved in the development of preeclampsia.[95,59,61] The aryl hydrocarbon receptor (AhR) mediates xenobiotic metabolism, dioxin and polycyclic aromatic hydrocarbon toxicity, and vascular development,[109] and AhR expression is elevated in preeclamptic pregnancy placentas.[111]

It is difficult to study each of these cigarette smoke components as independent exposures, as they are typically found as mixtures produced through combustion. Investigating gene by smoking interactions may provide some insight into the inverse smoking and preeclampsia relationship. We aim to assess multiplicative interactions between maternal smoking and genetic variants in pathways involved in response to cigarette smoke components on risk of preeclampsia. Observing differential associations between smoking and preeclampsia by genotype could provide evidence for why some individuals experience a protective effect of smoking and potentially identify mechanistic targets for future research.

Methods

We performed a nested case-control study within the Norwegian Mother and Child Cohort Study (MoBa),[222] a large prospective birth cohort of pregnant women and their offspring recruited throughout Norway from 1999 to 2008 (N=112,908 pregnancies), which has been previously described elsewhere.[222] Participants completed two prenatal questionnaires about their health and environment. Survey completion rate was 91% for the early pregnancy questionnaire and 83% for the late pregnancy questionnaire.[222] Maternal blood was collected at

the first ultrasound appointment and cord (child) blood was collected at birth. Maternal blood was received from 89% of participants and child blood from 81% of children in the cohort.[222] DNA was extracted at the time of collection before being stored at the MoBa Biobank.

Women provided informed consent prior to participation in MoBa and the study was approved by the Institutional Review Board of the Norwegian Institute of Public Health. The current study was approved by the Institutional Review Boards of the Norwegian Institute for Public Health and the University of North Carolina at Chapel Hill.

Outcome Assessment

Preeclampsia was defined by the American College of Obstetrics and Gynecologists (ACOG),[27] which specifies that both the following be present:

6) Systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure of ≥ 90 mm Hg occurring after 20 weeks' gestation in a woman whose blood pressure has been previously normal, and
7) Proteinuria, with excretion of ≥ 0.3 g of protein in a 24-hour urine specimen or as measured by 1+ on urine dipstick.

Preeclampsia information from MoBa was obtained through linkage with the Medical Birth Registry of Norway (MBRN).[247] All registered preeclampsia cases (n=3500) and a random sample of 2000 pregnancies registered as being unaffected by preeclampsia were selected from the MoBa cohort to be verified by antenatal records through an independent validation study.[223] A 2013 revision of the ACOG definition also included clinical symptoms, however, we used the above criteria, current at the time of the validation study. We also considered as "true cases" any case with an ICD-10 code of severe PE (O14.1) or HELLP syndrome (O14.2) and delivery < 37 weeks, or ICD-10 code of eclampsia (O15), which are routinely validated by the MBRN.

For the current study, we included all women from the validation study with a singleton pregnancy who conceived spontaneously, were verified cases or controls, returned both the early and late pregnancy questionnaires, and had no history of chronic hypertension. In total, 2682 preeclampsia case samples (1564 maternal and 1118 fetal blood samples) and 1967 non-preeclampsia control samples (999 maternal and 968 fetal samples) met inclusion criteria and were genotyped.

Maternal Smoking

We created a dichotomous smoking variable to indicate self-report of any smoking in gestational weeks 11 through 20. We selected this time period to specifically capture smoking during the window in which maternal blood flow perfusion into the intervillous space occurs and trophoblasts move to an invasive state completing spiral artery remodeling.[228] Women were surveyed twice during pregnancy about smoking habits. In the early survey (13-17 weeks'), women were asked whether they had smoked prior to pregnancy, whether they currently smoked, and if so, how many cigarettes per day or week. If they were not current smokers, they were also asked if they had stopped smoking after becoming pregnant, and if so, at what gestational age. In the late questionnaire (~30 weeks'), women were asked whether they currently smoked and if so, how much. They were also asked if they had quit smoking during pregnancy, and if so, at what gestational age they stopped.

For our primary smoking variable during the window of 11 to 20 weeks' gestation, we used smoking information from the early pregnancy questionnaire unless it was missing or ambiguous, in which case we supplemented with information from the late pregnancy questionnaire. A woman was considered to be a non-smoker if she indicated she had never smoked or that she did not currently smoke, and we had no other evidence of smoking after 10 weeks of gestation. A woman was considered a smoker in weeks 11 to 20 if she indicated that she currently smoked or quit after 10 weeks' gestation on either survey. To determine the week

in which a woman quit smoking, we obtained the latest quit week reported on either the early or late pregnancy questionnaire. If a woman was missing smoking information on the early questionnaire and indicated being a daily smoker on the late questionnaire, we assumed she had also been smoking in weeks 11 to 20. To determine smoking quantity, we categorized smoking quantity as number of cigarettes reported daily or weekly. In a sensitivity analysis, we separately considered women who smoked during the entire 11-20 week window, and women who quit during the 11-20 week window.

SNP Selection

For this study, 124 genes involved in response to cigarette smoke components were identified from 8 canonical pathways (Table S2), which included: 1) endothelial nitric oxide synthase (eNOS) signaling pathway, 2) heme degradation, 3) hypoxia-inducible factor 1-alpha (HIF1A), 4) xenobiotic metabolism, 5) aryl hydrocarbon receptor signaling, 6) glutathione-mediated detoxification, 7) nicotine degradation II, and 8) nicotine degradation III. A total of 1,915 SNPs (MAF $\geq$ 10%) were selected and analyzed for this study, using a 10kb upstream and downstream margin around the transcription start and end sites of each gene.

Genotyping and Quality Control

SNPs were genotyped by the UNC Mammalian Genotyping Core using the HumanCoreExome+ array from Illumina (Illumina, Inc., San Diego, CA). Samples and SNPs were examined using PLINK 1.07 (http://pngu.mgh.harvard.edu/purcell/plink) for quality control. SNPs were excluded if the missing rate exceeded 5%, there was substantial deviation from Hardy-Weinberg Equilibrium ($p < 1x10^{-3}$) or minor allele frequency was <10%. Known genotype and DNA replicates were included on each plate and exhibited high genotyping quality. All subject-specific call rates were acceptable (minimum 97.2%). Sex-specific markers were

inspected and parent-child relatedness and inbreeding within the cohort was assessed by identity by descent.

Quantile-quantile plots and calculation of genomic control lambda[226] (lambda mom=1.01, lambda child=1.03) indicated no systematic test statistic inflation, unidentified relationships, or cryptic admixture. Outliers for each of the first three principal components >3 standard deviations from the mean were excluded. The final analysis sample (n=4514 total samples) included dyads with both mother and child genotype data as well as incomplete dyads with only mother or child genotype data (n=2596 preeclampsia case samples [1063 mother/child pairs, 450 mother only, 20 child only], n=1918 control samples [925 mother/child pairs, 45 mother only, 23 child only]).

Statistical Analysis

The inverse relationship between smoking and preeclampsia in the MoBa cohort has been reported elsewhere (OR=0.60, 95% CI 0.48 – 0.75 for second-trimester active smoking).[75] The primary goal of this analysis was to evaluate whether genetic variants in pathways involved in response to cigarette smoke components modify the maternal smoking-preeclampsia relationship.

To address this goal, we extended the case-mother control-mother log-linear modeling approach proposed by Shi et al.,[213] which simultaneously accommodates maternal and fetal genotypes, as well as maternal smoking, covariates, and maternal smoking-genotype interaction (both mother and child). This method uses Poisson regression to model expected counts of each possible genetic mating type combination (the set of genotypes in the parents) under the assumption of Mendelian inheritance as follows:

$$\ln[E(N_{mcde})] = \theta_{mc} + \delta d + \gamma I_e + \sigma d I_e + \alpha_1 d I_{m=1} + \alpha_2 d I_{m=2} + \beta_1 d I_{c=1} + \beta_2 d I_{c=2} + \omega d I_e \times G$$

Where $E(N_{mcde})$ is the expected value of the counts of families with each of maternal genotypes, child genotypes, case or control status, and smoking status; d is case or control status where

$d = 1$ for a case and $d = 0$ for a control; $I_{(e=1)}$ is an indicator for maternal smoking; and *G* is the number of copies of the variant allele that the mother carries (when assessing maternal interaction) or that the child carries (when assessing child interaction). The θ_{mc} parameters allow flexibility of the control-mother distribution (avoiding the need to assume Hardy-Weinberg for the source population) and ensure that the parental genotype distribution is only constrained by the family relationships.

LEM software[236] was used to fit these models. The expectation maximization algorithm was used to incorporate dyads with missing genotypes. The proportion of missing smoking data was low (<5%), so observations missing smoking data were excluded from analysis. Likelihood ratio tests comparing reduced models without the interaction term with the model including an interaction term were performed to determine p-values for interaction for both maternal-genotype interaction and child-genotype interaction, each adjusted for the other genotype.

Point estimates and 95% confidence intervals for relative risks were calculated for risk of preeclampsia for maternal smokers compared to non-smokers during the window of 11 to 20 weeks of gestation, stratified by both maternal genotype and child genotype.

We calculated Bonferroni corrected p-values to account for multiple comparisons ($P < 2.6 \times 10^{-5}$). Interaction tests generally have low power to reject homogeneity,[240] thus, we also report results for a relatively generous cutpoint that less completely acknowledges issues of multiple testing ($P < 0.001$).

Due to the potential for exposure-related population stratification bias within family-based studies, as reported by Shi et al.,[262] we also performed a sensitivity analysis in which we extended the model to include smoking by mating type interaction parameters to allow exposure status to vary by mating type. This model is as follows:

$$\ln[E(N_{mcde})] = \theta_{mc} + \theta_{mc}I_e + \delta d + \gamma I_e + \sigma d I_e + \alpha_1 d I_{m=1} + \alpha_2 d I_{m=2} + \beta_1 d I_{c=1} + \beta_2 d I_{c=2} + \omega d I_e \times G$$

in which $\theta_{mc}I_e$ is now a correction to the mating type parameters where the mother smoked.

Results

The final analysis sample consisted of 4,288 individual samples for 1,888 complete mother-child dyads, 471 dyads with only maternal genotype data (429 cases), and 41 dyads with only child genotype data (19 cases) (n=2,400 pregnancies) (Table 13). Of the 2,526 pregnancies for which we had genotype information, 126 (5.0%) were missing information on smoking status during gestational weeks 11 to 20. Distribution of covariates did not differ between all pregnancies and those missing smoking data (data not shown). Of the 2,400 pregnancies for which we had smoking information, 214 (8.9%) smoked during weeks 11 to 20. Fewer women with preeclampsia smoked during weeks 11 to 20 than those without (8.2% vs 10.0%). Smoking was light overall in this population. More women with preeclampsia quit smoking during the 11 to 20 week window than those without preeclampsia (33% vs 25%). Women with preeclampsia were more often nulliparous than those without (66% vs 41%) and tended to have higher body mass index. Infants born to women with preeclampsia were more often preterm (21% vs 3%) or small for gestational age (23% vs 8%).

In analyses of interactions, none of our SNPs met our Bonferroni-corrected threshold ($P < 2.6 \times 10^{-5}$). We present maternal interaction results for the three SNPs with p-values for interaction <0.001 (Table 14). We found no SNPs meeting this threshold for child genotype interaction. Risk ratios for the association between smoking and preeclampsia are stratified by genotype (Table 14).

Among women who had the C allele of rs3765692 (*TP73*), risk of preeclampsia was lower for smokers compared to non-smokers (RR among *CT*: 0.53, 95% CI 0.36 – 0.79; RR among *CC*: 0.28, 95% CI 0.14, 0.58) compared to a null association among those with 0 copies of the variant allele (RR among *TT*: 1.00, 95% CI 0.74 – 1.36). The pattern was similar for rs10770343 (*PIK3C2G*), where among women with the C allele, smokers had a reduced risk of preeclampsia (RR among *CA*: 0.63, 95% CI 0.45 – 0.88; RR among *CC*: 0.37, 95% CI 0.21 –

0.65) compared to a null association among those without the variant allele (RR among *AA*: 1.07 95% CI 0.78, 1.48). The pattern was opposite for rs2278361 (*APAF1*), in which smokers had an increasing risk of preeclampsia with increasing number of variant alleles. For rs2278361, a null association of smoking and preeclampsia was found among heterozygotes (RR for *CT*: 1.04, 95% CI 0.76 – 1.41), while associations among the *TT* and *CC* genotypes straddled each side of the null (RR among *TT:* 0.62, 95% CI 0.44 – 0.86; RR among *CC:* 1.74, 95% CI 1.05 – 2.88). Results are presented graphically in Figure 15.

Results of the sensitivity analysis allowing for exposure-related population stratification are similar, although given how rare smoking is, we had a number of mating types for which there were no smokers, and thus we could not fully evaluate the extent to which population-stratification by exposure may influence our findings. Nonetheless, although estimates were less precise, the same set of SNPs had the lowest p-values for smoking-SNP interactions in mothers, and risk ratios were similar in magnitude and direction for those SNPs (Table S13a). Two additional smoking-SNP interactions in children emerged as potentially noteworthy, but estimates were unstable due to sparse data (Table S13b). Results of our sensitivity analysis of smoking cessation revealed some differences among groups. Table S14 compares risk ratios and p-values for interaction for groups in which smokers consisted of 1) only those who quit during weeks 11-20, 2) only those who did not quit during weeks 11-20, and 3) all smokers during weeks 11-20 (our primary analysis). Risk ratios are presented per-minor allele to more easily compare the differences in associations among groups. As in the primary analysis of the group consisting of all smokers, the direction of the associations remained the same. Risk of preeclampsia for smokers compared to non-smokers decreased with each additional minor allele for rs3765692 (*TP73*) and rs10770343 (*PIK3C2G*), and increased for rs2278361 (*APAF1*). Although the direction of association was the same for all groups, associations were further from the null for the group containing smokers that did not quit smoking during the 11 to 20 week window.

Discussion

The inverse relationship between smoking and preeclampsia is well established, though the reason for this association is still unknown.[6] This study employed a novel approach in an effort to determine if there are genetic subgroups in which there are differences in the association between smoking and preeclampsia is due to variation in genes involved in smoking-related pathways. We assessed possible gene-by-maternal-smoking interaction for 1915 candidate SNPs representing 8 canonical pathways related to biologic responses to the tobacco combustion products.

We found little evidence for interaction between maternal smoking during the time of spiral artery remodeling and genetic variants in pathways related to biologic responses to cigarette smoke components. While not statistically significant, for our most noteworthy SNPs (p-int < 0.001), rs3765692 in *TP73*, rs1077343 in *PIK3C2G,* and rs2278361 in *APAF1* we showed a differential effect of smoking on preeclampsia by genotype. The patterns for these SNPs suggest a biological mechanism related to effects of the exposure that are mediated through that gene may partially explain the association.

These SNPs should be investigated for more targeted epidemiologic and mechanistic research as their function is currently unknown but each has some emerging evidence for biological plausibility. The P3IK pathway (in which *PIK3C2G* is included) has recently been reported to affect placental maternal-fetal resource allocation. [263,264] Aryl hydrocarbon receptor activation represses *TP73* and *APAF1*, resulting in pro-proliferative, anti-apoptotic effects.[265] Both have biologic plausibility in the pathogenesis of preeclampsia.

Studies of gene (*TNFα, CYP1A1, GSTT1)* by smoking interactions have focused on other reproductive outcomes such as preterm birth[215] and orofacial clefts, [191214] while the epidemiologic study of gene by environment interaction for preeclampsia has been limited to the study of the *MTHFR* gene and folate.[191] We selected a comprehensive list of SNPs from genes in pathways involved in response to cigarette smoke components based on biological

plausibility[111,266,92,93] and prior interactions with smoking in other conditions such *CYP1A1, CYP1B1, GSTM1,* and *GSTT1* with lung,[267] head and neck,[268,269] and colorectal[270] cancer.[266]

In addition to approaching the question of how smoking affects preeclampsia in a new way by assessing gene by environment interactions, this study aimed to address several common methodological problems in the study of preeclampsia. Misclassification due to misdiagnosis and reporting is often a problem with preeclampsia.[271] Preeclampsia was verified by antenatal medical records and hospital diagnostic codes through medical record validation.[223] All cases of preeclampsia in our study were verified as having preeclampsia and all controls were verified as being free of preeclampsia. The time window for maternal smoking exposure may be an important determinant of risk, and classification of smoking by time period has been a challenge in studies of preeclampsia. Biologically, smoking during early pregnancy may influence outcomes differently than in late pregnancy because of differences in the placenta, mother, and fetus. Because of the negative effects that smoking can have on the fetus and the associated cultural pressures, many women stop or reduce their smoking when they find out they are pregnant. We had two different prenatal surveys, and both asked about weeks of smoking, so we were able to classify by several different time points during pregnancy. We selected one specific interval (gestational weeks 11 to 20) as the most biologically relevant exposure time for our primary analyses.

We took several approaches to improve our study power, which is a common difficulty in gene by environment studies. Use of a case-mother control-mother design incorporating the expectation maximization algorithm to account for missing data improved power by allowing dyads where only one member was genotyped to contribute fully. We modeled the gene by environment interaction as a log-additive interaction term to improve power, although that may not be the most appropriate model fit. Additionally, we chose to limit our analysis to SNPs with both a plausible biological relationship with effects of smoking or known interaction with smoking, and also limited to minor allele frequencies greater than 10%. We also selected SNPs

using a relatively small 10kb upstream and downstream margin around the transcription start and end sites for each gene to ensure that we captured both the entire coding region of the gene and proximate regulatory regions, while constraining the number of SNPs we were studying to preserve our study power. This may have limited our ability to detect distal intergenic variants that may alter gene expression through their role within the regulatory regions.[258] Despite these strategies, we are still underpowered to assess gene by smoking interactions for the number of SNPs in our study.

Our method accounted for family structure and included both mother and child DNA, each adjusting for the other. Given the suspected pathophysiology underlying preeclampsia, it is biologically plausible that both maternal and child genotypes contribute to this pregnancy complication. A few other genetic studies of preeclampsia have addressed child genotype[260,124,194] but have not simultaneously modeled both mother and child genotype while addressing gene by smoking interactions.

Although we were able to account for confounding by both maternal and child genotype in our study, we may have inadequately addressed other types of confounding in our analysis. Maternal age and parity are associated with both smoking and preeclampsia in our cohort, and may both be potential confounding variables in the smoking-preeclampsia association. The mating type parameter makes inclusion of other covariates challenging due to non-positivity within mating types and maternal age and parity were not included in our model. Our model may also be vulnerable to exposure-related population stratification, which has been shown to bias results if both smoking exposure rates and mating type frequencies differ among subpopulations.[262] We controlled for exposure-related population stratification bias in an sensitivity analysis by including additional family-based exposure parameters (Table S13), and results were substantially unchanged, but the small number of smokers in our dataset limits our ability to make inferences from models that include these additional parameters.

It is a challenge to ascertain smoking exposure. Our study used self-reported smoking status from maternal questionnaires. Smoking status may be underreported due to social stigma, particularly among pregnant women where the relationship between smoking and poor birth outcomes is known.[272] A recent validation study of self-reported smoking and plasma cotinine indicated that reported smoking on the MoBa questionnaire is a reasonably valid marker of tobacco exposure (Sensitivity=82%, Specificity=99%),[227] however, even this degree of measurement error may have biased our results. Differential false reporting of smoking status may vary by preeclampsia case status, which could have influenced our ability to detect any effect measure modification, however, in most cases, ascertainment of smoking status came before preeclampsia diagnosis so this is unlikely an issue.

Additionally, pregnancy is a time period in which many women quit smoking after learning they are pregnant or are actively trying to quit smoking throughout pregnancy. In our study 19.5% of women reported smoking in the first trimester and 5.9% in the third trimester. When assessing SNP by smoking interaction by time period, estimates were similar, but attenuated toward the null in the first trimester (data not shown). Although we selected a specific, biologically relevant time period for smoking, 63 of 214 (29%) of women who smoked during weeks' 11-20 reported quitting in a week during that period. To address whether quitting specifically influenced our results, we conducted a sensitivity analysis to see if people who reported quitting were different than those who continued smoking. The direction of association was the same for all groups, but associations were stronger and more significant for the group of smokers who did not quit during weeks' 11-20 compared to those who reported that they did quit, indicating there may be a dose-response effect which should be investigated further (Table S14).

Although smoking quantity may be important, we measured smoking dichotomously as any smoking during the time period compared to no smoking. Although we did have information on the average number of cigarettes smoked per day/week, overall smoking intensity was light

in this population, and stratification, along with the infrequency of smoking overall, presented sparse cells when stratified by case and mating type parameters. This low dose of smoking may have also influenced our ability to identify relevant gene-smoking interactions.

Overall, we found little evidence of multiplicative interaction between SNPs in smoking response pathways and maternal smoking during spiral artery remodeling. The case-mother control-mother design as well as environmental interaction analysis may be useful methods to include in the toolkit for the genetic study of pregnancy outcomes, particularly preeclampsia. However, very large studies will be needed in order to be adequately powered to examine exposures by mating type, genotype, case-status and dose.

Tables

Table 13. Characteristics of participants by case and control status (n=2,400 pregnancies)

	Preeclampsia Cases (N = 1451)		Controls (N = 949)	
	No.	%	No.	%
Maternal Age[a]				
≤ 20 years	45	3.1	14	1.5
21 – 30 years	834	57.5	462	48.7
31 – 40 years	551	38.0	468	49.3
≥ 41 years	20	1.4	5	0.5
Maternal Education[a]				
< High School	127	9.3	69	7.8
High School Graduate	431	31.6	255	28.7
University Degree	804	59.0	565	63.6
Body Mass Index (kg/m^2)[a]				
Underweight (<18.5)	23	1.6	32	3.5
Normal weight (18.5-24.9)	686	48.8	612	66.6
Overweight (25.0-29.9)	437	31.1	181	19.7
Obese (30.0+)	259	18.4	94	10.2
Any Maternal Smoking (11-20 weeks)	119	8.2	95	10.0
Smokers who quit during 11-20 wks[b]	39	32.8	24	25.3
Smoking Intensity[a, b]				
0-2 cigarettes/day	34	36.5	37	46.3
3-5 cigarettes/day	30	31.3	19	23.8
6-10 cigarettes/day	28	29.2	19	23.8
>10 cigarettes/day	3	6.3	5	3.1
Nulliparous[a]	954	65.8	392	41.3
Preterm (< 37 weeks)[a]	305	21.2	28	3.0
Small for gestational age (SGA) (< 10th percentile)[a, c]	323	22.5	74	7.8
Severe preeclampsia	285	19.7	--	--
Birthweight (mean grams (SD))[a]	3173.4 (820.3)		3676.74 (512.4)	

[a]Missing observations for each covariate: maternal age (1), maternal education (149), body mass index (76), smoking intensity (39), parity (1), preterm birth (11), small for gestational age (16), birthweight (1).
[b] Percentage calculated as proportion of women who smoked during 11-20 week window.
[c] Population percentiles derived from Norwegian distribution, eSnurra Norway.

Table 14. Relative risks for the association of smoking and preeclampsia for all maternal and child smoking-SNP interactions with interaction $p < 0.001$, stratified by genotype.

						Mother		Child	
Marker[a]	**Chr**	**Position**	**MAF**	**Gene**	**Genotype**	**RR (95% CI)**	**P-Interaction**	**RR (95% CI)**	**P-Interaction**
rs3765692	1	3584771	0.22	*TP73*	*TT*	1.00 (0.74, 1.36)		0.91 (0.66, 1.24)	
					CT	0.53 (0.36, 0.79)	5.9×10^{-4}	0.66 (0.45, 0.97)	0.09
					CC	0.28 (0.14, 0.58)		0.48 (0.24, 0.96)	
rs10770343	12	18414253	0.31	*PIK3C2G*	*AA*	1.07 (0.78, 1.48)		1.02 (0.73, 1.43)	
					CA	0.63 (0.45, 0.88)	5.9×10^{-4}	0.67 (0.48, 0.93)	0.02
					CC	0.37 (0.21, 0.65)		0.44 (0.24, 0.80)	
rs2278361	12	99043207	0.21	*APAF1*	*TT*	0.62 (0.44, 0.86)		0.66 (0.47, 0.92)	
					CT	1.04 (0.76, 1.41)	7.5×10^{-4}	0.99 (0.72, 1.37)	0.02
					CC	1.74 (1.05, 2.88)		1.49 (0.84, 2.65)	

Figures

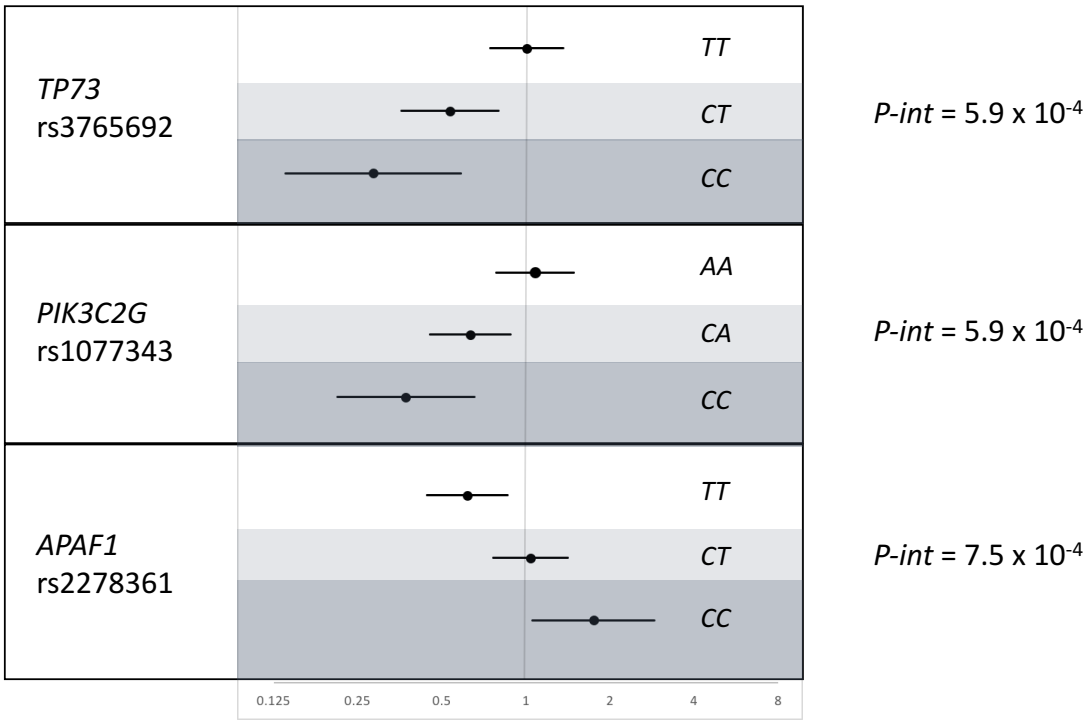


Figure 15. Risk ratios for the association of smoking on preeclampsia stratified by genotype for the top three SNP-maternal genotype interactions.

Additional Results

Models were run for aim 2 (as in aim 1) for the association of SNP and preeclampsia for the additional 397 smoking detoxification SNPs. There were no additional significant findings. Analyses were run separately restricting to 1) nulliparous women, and 2) overweight women with preeclampsia diagnosed after 34 weeks' gestation. Results were not substantially different than in the full cohort.

CHAPTER VI. DISCUSSION

A. Summary of Findings

This study was motivated by the well-established inverse relationship between smoking and preeclampsia, the reasons for which remain unknown.[6] Although methodological reasons for this inverse association have been suggested,[261] this study aimed to investigate potential biological reasons for the association.

The first aim was to determine if variants in carbon monoxide and nitric oxide signaling genes are associated with preeclampsia, both overall, and among preeclampsia subtypes. This aim investigated genes that may mimic smoking as a proxy for exposure. For example, it is difficult to measure nitric oxide in the placental tissue and uterine arteries, however, we can measure genotype within the nitric oxide synthase gene, which has a direct effect on synthesis of nitric oxide in the body. This genetic pathway is of interest because there are pharmaceuticals that directly operate through the action of nitric oxide that could potentially be used as a therapeutic treatment for preeclampsia. An association between a gene in these pathways and preeclampsia could provide support for specific ways in which components of cigarette smoke influence the development of preeclampsia.

In our aim 1 analysis, assessing associations of SNPs within carbon monoxide and nitric oxide signaling pathways with preeclampsia, we found one SNP meeting our criterion of *Q*-value <0.05, a child variant, rs12547243 in adenylate cyclase 8 (*ADCY8*). We also saw some suggestive maternal associations of SNPs in *PDE1C* for preeclampsia accompanied by early delivery and preeclampsia accompanied by small for gestational age.

Biologically, these associations are interesting and relevant. *ADCY8* encodes adenylate cyclase 8 which catalyzes the conversion of adenosine-5'-triphosphate (ATP) to 3', 5'-adenosine monophosphate (cAMP).[249] *PDE1C* encodes calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1C and catalyzes conversion of cAMP and cGMP to their 5-prime-monophosphates (AMP and GMP).[251] Both are part of the canonical pathway for cellular effects of sildenafil ("Viagra"), a phosphodiesterase inhibitor that operates through nitric oxide signaling to increase vasodilation,[252] and mouse studies have demonstrated resolution of the preeclampsia phenotype with the administration of sildenafil.[253,254]

The second aim was to determine if genetic variants in carbon monoxide and nitric oxide signaling genes, as well as smoking detoxification genes modify the effect of smoking on preeclampsia. Our hypothesis was that the smoking-preeclampsia relationship could be affected by detoxification of components of cigarette smoke (e.g. nicotine, aryl hydrocarbons) or through the additional endogenous production of byproducts (NO/CO). Although there is a strong inverse association between smoking and preeclampsia, there is individual variation in those who experience a reduced risk of preeclampsia in the presence of smoking; many women who smoke still get preeclampsia. Biologically, it is possible that smoking activates a gene that induces a response that reduces susceptibility to preeclampsia. For example, there are genes that are activated by smoking such as the aryl hydrocarbon receptor gene (*AHR*) and members of the *GST* and *CYP* families of genes. The encoded proteins are hypothesized to protect the body against the harmful effects of smoking, however, may also have more general protective properties such as reducing oxidative stress and inflammation and increasing angiogenesis. Although the specific gene may not directly affect preeclampsia, activation in the presence of smoking may induce a response. Given the strong effect of smoking on preeclampsia, genes that are affected by smoking may modify the overall association between smoking and preeclampsia. Observing differential associations between smoking and preeclampsia by

genotype could provide evidence for why some individuals experience a protective effect of smoking and potentially identify mechanistic targets for future research.

In aim 2, we found limited evidence for multiplicative SNP by smoking interaction after correction for multiple testing but saw a few potentially interesting findings. We saw decreasing risk of smoking on preeclampsia with increasing number of maternal variant alleles for rs3765692 (*TP73*) and rs10770343 (*PIK3C2G)* and an increasing risk of smoking on preeclampsia with increasing number of maternal variant alleles for rs2278361 (*APAF1*).

The P3IK pathway (in which *PIK3C2G* is included) has recently been determined to affect placental maternal-fetal resource allocation.[263,264] And the aryl hydrocarbon receptor activation represses *TP73* and *APAF1*, resulting in pro-proliferative, anti-apoptotic function,[265] all of which may play a role in preeclampsia. One could argue, however, that all of these genes were selected for their biological relevance to preeclampsia, and thus, we might expect to see some interesting biological patterns among our results.

B. Strengths and Limitations

Strengths

This study is innovative in that it is one of the few to tackle the potential reasons for the inverse smoking-preeclampsia association and explore potential biological mechanisms using epidemiologic data. In particular, there are very few studies of gene by environment interactions that have been conducted for preeclampsia and reproductive outcomes in general, and to our knowledge, this is the first of gene by smoking interaction with preeclampsia.

A major advantage of this study is that it is nested within a cohort whose target population is all pregnant women in Norway. Selection from a well-defined cohort improves both internal and external validity. Because of the structure of healthcare in Norway, limited access to prenatal care does not impede recruitment as it would in other studies in which women are

recruited through prenatal care providers. This can increase generalizability to a broader population, potentially making the study population more comparable to the target population.

Prior genetic studies of preeclampsia have been limited by power. Our study is one of the largest genetic studies of preeclampsia to date, which allowed us to address both mother and child genotype. Our study design takes advantage of family-based constraints and uses the expectation maximization algorithm to account for missing data and improve power. Although there are serious criticisms of candidate genes studies,[273] this study, though not whole-genome scale, aimed to improve upon prior candidate gene studies by including more genes and SNPs while continuing to apply a hypothesis-driven approach to maintain statistical power.

Given the suspected pathophysiology underlying preeclampsia, it is biologically plausible that both maternal and child genotypes contribute to this pregnancy complication. Our dyad analysis accounted for family structure by including both mother and child DNA, each adjusting for the other by characterizing associations with mating types. Few other genetic studies of preeclampsia have addressed child genotype[124,194,260], and even fewer have simultaneously modeled both mother and child genotype.[274]

While addressing limitations of previous studies by accounting for both maternal and child genetic components, this study also took on the challenges of misclassification and heterogeneity of the preeclampsia phenotype. Preeclampsia was verified by antenatal medical records and hospital diagnostic codes through medical record validation.[223] All cases of preeclampsia in our study were verified as having preeclampsia and all controls were verified as being free of preeclampsia. Using this information, we were also able to classify cases into preeclampsia subtypes that may have differing underlying etiologies an aspect of preeclampsia that has not been well-explored in genetic studies.

In addition to addressing some of the challenges of defining the preeclampsia outcome, we also addressed challenges of defining maternal smoking exposure. We did so by selecting a specific etiologic window during which aberrant placentation may affect development of

preeclampsia. For our primary analysis, we combined information from multiple questions about smoking timing and behavior to define smoking as any smoking during gestational weeks' 11 to 20. We selected this time period to specifically capture smoking during the period in which maternal blood flow perfusion into the intervillous space increases and trophoblasts move to an invasive state completing spiral artery remodeling.[228]

Limitations

Candidate gene studies have had limited success in identifying causal variants in the past,[275] and this study is no exception. Although we use prior evidence of biological significance to inform this hypothesis-driven study, genome wide association studies are typically a more efficient and thorough way of identifying potential associations. Prior GWA studies of preeclampsia, however, have suffered from insufficient sample size and heterogeneity of ancestry. Our study used a hybrid, hypothesis-driven approach to examine SNPs in whole pathways while still reducing the number of statistical tests compared to a GWAS.

Our gene by environment analysis was limited by statistical power. In the gene by environment analysis, we chose to limit our analysis to SNPs with plausible biological significance for their relationship with smoking, and with minor allele frequencies greater than 10%. However, despite doing so, we are still underpowered to assess gene by smoking interactions for the number of SNPs in our study, particularly when mating types were allowed to differ by exposure status, and when smoking was considered by continuity across the exposure window.

Our study assessed multiplicative SNP by smoking interactions. However, additive interactions may deliver a more easily translatable public health message insofar as they describe the absolute risk difference among subgroups.[152] We were only able to estimate multiplicative interactions, but it is possible that important additive interaction exists.

Development of good methods for assessing additive interactions based on a case-parent/control-parent approach is a matter for future work.

We attempted to replicate our findings within UK data from the InterPregGen Consortium, a consortium of studies of genetics of preeclampsia. Although our ability to include mother-child pairs is innovative, most available preeclampsia data sources do not include paired mother-child dyads. Our replication approach analyzed both mother and child genetic data, but the data were unpaired, and thus, may not be comparable.

Although our analytic method is innovative, improves power, and allows us to account for mother and child genetic contributions, this study may be subject to several types of confounding, misclassification, and selection biases.

Confounding

This study controlled for paired mother-child genotypes, which is likely the biggest confounder However, a limitation of this study design is the challenge of controlling for external confounders. After accounting for maternal and child genotype, in the SNP-preeclampsia relationship, there is little confounding since there is little that directly affects genotype. However, ancestry does affect genotype and given that inference under this design relies on comparison of case-mother pairs with control-mother pairs, there is still room for bias due to population stratification. Moreover, because we are potentially interested in the joint effect of both SNP (or related biological mechanisms) and maternal smoking, we would want to control for confounding in the relationship of the factor defining the subgroups (smoking) as well.[276] Maternal age and parity are associated with both smoking and preeclampsia in our cohort, and both are potential confounding variables in the smoking-preeclampsia association. Maternal age and parity were not included in the model, however, age is not associated with preeclampsia after adjusting for nulliparity or related to allele frequency of the SNPs in our study so confounding due to these variables is unlikely an issue.

Population stratification bias, resulting from systematic differences in allele frequencies by population substructure may lead to spurious associations. In a typical logistic regression analysis, principal components are included as covariates in the model to adjust for population substructure; as with other covariates, ability to include eigenstrata is limited as well. We addressed this limitation by assessing population stratification in the quality control process to determine if ancestral homogeneity was a plausible assumption within this Norwegian population. We compared our population with the 1000 Genomes reference populations and excluded outlier observations in a sensitivity analysis.

Our model may also be vulnerable to exposure-related population stratification, which has been shown to bias results if both smoking exposure and mating type frequencies differ among subpopulations.[262] We controlled for exposure-related population stratification bias in a sensitivity analysis by including additional family-based exposure parameters (Table S13), and results were substantially unchanged.

Classification of Smoking

Our study used self-reported smoking status from maternal questionnaire. Use of self-reported data, as well as the changing smoking behaviors of women throughout the course of pregnancy may result in misclassification bias. Smoking status may be underreported due to social stigma, particularly among pregnant women where the relationship between smoking and poor birth outcomes are known.[272] A recent validation study of self-reported smoking and plasma cotinine indicated that reported smoking on the MoBa questionnaire is a fairly valid marker of tobacco exposure (Sensitivity=82%, Specificity=99%).[227] However, this degree of measurement error may have biased our results.

Additionally, pregnancy is a time period in which many women quit smoking after learning they are pregnant or are actively trying to quit smoking throughout pregnancy. In our study 19.5% of women reported smoking in the first trimester and 5.9% in the third trimester.

When assessing SNP by smoking interaction by time period, estimates were similar, but attenuated toward the null in the first trimester (Fig S1). Although we selected a specific, biologically relevant time period for smoking, 63 of 214 (29%) of women who smoked during weeks' 11-20 reported quitting in a week during that period. To address whether quitting may have influenced our results, we conducted a sensitivity analysis to see if people who reported quitting were different than those who continued smoking. The direction of association was the same for all groups, but associations were stronger and more significant for the group of smokers who did not report quitting during weeks' 11-20 compared to those who did quit, indicating there may be a dose-response effect that should be investigated further (Table S14).

Although smoking quantity may be important, we analyzed smoking dichotomously (ever versus never in that period). Although we did have information on smoking intensity of number of cigarettes smoked we were unable to assess smoking dose, as our data were already sparse before stratifying by smoking. Table S15 shows the data table for one of our lead SNPs demonstrating that cell counts are already sparse without further stratifying by smoking intensity. Additionally, this is a population of relatively light smokers; even the heaviest smokers were smoking only approximately half a pack of cigarettes a day.

Selection Factors

This study is also open to several forms of selection bias. Possible selection bias may occur due to refusal to participate. MoBa is a long-term prospective cohort study, and the criteria to participate are somewhat stringent; participants must agree that MoBa can access health registries for purposes within the broad aim of the study and they agree to give biospecimens from themselves and their infant. The participation rate is low, at approximately 45%.[221] When MoBa study participants were compared with the population of all births in Norway during the same time period, prevalence of preeclampsia was similar, but women in MoBa were less likely to be young (<25 years), live alone, smoke, and have more than two

births.[277] When associations between covariates and birth outcomes in the two populations were compared, however, there were few differences (e.g. the ratio of adjusted odds ratios between smoking and low birthweight and between parity and preeclampsia were approximately 1.0).[277]

A consistent problem in the study of preeclampsia is the problem of left truncation, which may result in another form of selection bias. Lisonkova and Joseph posit that left truncation bias due to differential rates of early pregnancy loss among smokers is a plausible explanation for the inverse relationship between smoking and preeclampsia.[261] They make the following assumptions in their simulation model: 1) abnormal placentation leads to preeclampsia, 2) similar rates of abnormal placentation occur in smokers and non-smokers, 3) higher rates of early pregnancy loss occur among smokers versus non-smokers, 4) higher rates of early pregnancy loss occur among women with abnormal placentation than without, and 5) the highest rates of early pregnancy loss occur among smokers with abnormal placentation. By varying rate ratios of early pregnancy loss due to smoking and early pregnancy loss due to abnormal placentation, they illustrate that the inverse association with smoking may be a result of bias due to left truncation.

Because women enrolled in MoBa between 13 and 17 weeks' gestation, we were unable to capture early pregnancy losses and whether there is differential loss by smoking status. Therefore, we are unable to exclude left truncation as a potential mechanism for the smoking-preeclampsia relationship. However, Kinlaw and colleagues[278] recently re-examined the assumptions of Lisonkova and Joseph, and the impact of varying these assumptions on the inverse association of smoking. Lisonkova assumed: 1) all fetuses surviving past 20 weeks of pregnancy with abnormal placentation progress to preeclampsia, 2) no fetuses surviving past 20 weeks of pregnancy without abnormal placentation progress to preeclampsia, and 3) maternal smoking has no effect on abnormal placentation.[278] Kinlaw and colleagues present scenarios in which there is much less bias due to differential early pregnancy loss when more realistic assumptions about the contributions of abnormal placentation on preeclampsia are applied.[277]

Though we did not find any evidence of gene by smoking interaction of SNPs within any of the GST genes in our study, there is evidence that the glutathione S-transferase P1b-1b genotype is found more often in women with early pregnancy loss, and particularly among those who smoked cigarettes[279] demonstrating a factor that might influence differential early loss by smoking and also affect placentation. It is possible that the inverse association between maternal smoking and preeclampsia is due to some combination of a biological effect of smoking on preeclampsia, some bias due to differential early pregnancy loss (or competing risk by preterm birth), and some additional unknown mechanism(s). Our results contribute to the growing literature of biologically interesting findings of the relationship between smoking and preeclampsia.

C. Directions for Future Research

The estimated heritability of preeclampsia is approximately 55%,[280] and thus it is surprising that this study and previous studies have not found any strong and replicated genetic signals. Because of evidence that there are genetic contributors to preeclampsia, questions of genetic influences on preeclampsia should continue to be investigated, but may require novel methods to address.

In our study, SNPs were selected using a 10kb upstream and downstream margin around the transcription start and end sites for each gene. Although we selected these margins to ensure that we captured the entire coding region of the gene and proximate regulatory regions, recent studies suggest that distal intergenic variants can alter gene expression through their role within the regulatory regions.[258] Although there were no previously found genome-wide associations found for any of our SNPs, a search with HaploReg[281] revealed that several of them (rs12547243, rs30593, and rs10770343) were found to alter the regulatory motifs of proteins encoded by genes not included in the present study (NF-kappaB, TATA, and NRSF,

respectively). GWAS can capture these intergenic variants by looking across the whole genome but will require large sample sizes to be adequately powered.

The pathophysiology of preeclampsia, and likely involvement of the placenta in its etiology, may be another reason that previous genetic studies of preeclampsia have not found any significant genetic signal. Our most noteworthy results are for child-genotype associations with preeclampsia. A few other genetic studies of preeclampsia have addressed child genotype,[124,194,260] but most have focused on the maternal component. While our analytic method begins to address some of the complexity of correlated mother and child genotype, it does not assess their interaction, which may be important and is a future area of interest to our study. It is plausible that there is a maternal genetic factor that affects a physiological response in the mother and a different fetal genetic factor that influences fetal programming and both must occur simultaneously during pregnancy to support optimal growth of the placenta and fetus. Likewise, maternal-fetal incompatibility can contribute to development of disease, such as that that is experienced with hemolytic disease of the newborn with mother and child having different Rhesus factor.[282] Our current analytic methods cannot handle inclusion of mother-child genetic interactions because it would over-parameterize our model, but future studies could incorporate mother-child interactions through the extension of our design with the use of methods like the maternal-fetal incompatibility test developed by Sinsheimer and colleagues, which allow one to estimate direct effects of both maternal and child variant alleles, as well as their interactions through determination of whether maternal and fetal alleles are the same or different at each locus.[283] These methods are just beginning to be applied on a genome-wide scale.[284]

Additionally, there is some evidence that paternally-inherited imprinted genes and partner-genotype discordance may be important in the placenta and in the development of preeclampsia in particular. [285,286] A fetal contribution to preeclampsia may, in part, occur through paternal genes. Skjærven and colleagues found that men born after a pregnancy complicated

by preeclampsia had an increased risk of fathering a preeclamptic pregnancy (OR=1.5, 95% CI: 1.3, 1.7)[287] and Lie and colleagues found that risk for preeclampsia is increased for women in which the pregnancy was fathered by a man who has previously fathered a preeclamptic pregnancy in a different woman (OR=1.8, 95% CI: 1.2, 2.6).[127] Although our mother-child dyad design will account for paternal contributions through inclusion of child genotype, other family-based approaches are necessary to detect contributions of paternal imprinting or paternal genetic discordance. Future analyses of preeclampsia should consider using information from the father where available, and new studies should consider inclusion of fathers at study development. Although they have their limitations, family-based studies are useful in the toolkit to study reproductive outcomes and can be expanded to incorporate paternal genotype and imprinting, a relevant area of study for preeclampsia. The triad approach also is fully protective against bias due to population stratification.

Heterogeneity of the preeclampsia phenotype can limit the ability to detect genetic effects. This heterogeneity may be due to both how preeclampsia is defined and classified, and those changing definitions over time, as well as differences among the underlying etiologies of different types of preeclampsia. Our study was one of the few that had a validated measure of the preeclampsia phenotype and assessed subtypes, but further refinement of the phenotype in future studies could improve our ability to detect associations and investigate biological mechanisms. Use of other methods such as growth mixture models that investigate blood pressure and proteinuria repeatedly over time could enable one to identify trajectories over time and provide insight into other ways of classifying preeclampsia that may have different etiology. These analyses could incorporate smoking and other environmental information; for example, Macdonald-Willis and colleagues did a latent class analysis of proteinuria and found that smoking was weakly associated with odds of proteinuria after 37 weeks' gestation.[288] Examination of other early pregnancy physiological factors instead of just a preeclampsia diagnosis could also provide insight into genetic contributions. Preeclampsia is diagnosed after

20 weeks' gestation, which is after the hypothesized mechanisms contributing to preeclampsia have already occurred. There are methods in development to detect precursors to preeclampsia, such as mapping of the placenta vasculature. [289,290] Use of such methods in future studies could provide a more refined phenotype that may reveal different associations than diagnosed preeclampsia.

The lack of replication in the literature is likely principally due to low power, and our study of gene by smoking interactions was particularly underpowered. Although it would provide less specific information and allow less flexibility in the model, we could improve our study power by using a 2 degree of freedom test that is a joint significance test for SNP and SNP by exposure.[291] However, a preferred approach is to pool across several large European birth cohorts, which would increase the availability of paired mother-child dyad data and increase statistical power. There are multi-study consortia forming with the objective of studying genetics of preeclampsia.[241] These collaborative efforts may provide the necessary power to more thoroughly address this perplexing condition.

Finally, it is plausible that any genetic effects of preeclampsia may not be purely a single SNP association but rather epigenetic or polygenic. There are already studies underway that are using some of these novel methods for the study of preeclampsia[161,292] and future studies should continue to explore potential genetic mechanisms beyond single SNP-outcome associations

D. Public Health Significance

Preeclampsia is a well-studied condition, though one in which there are still many unknowns about underlying mechanisms and etiology. We have known of the inverse association with maternal smoking for decades, yet the reason for this relationship is still

elusive. This study contributes to the small existing literature exploring the smoking-preeclampsia relationship and the growing literature on genetics and preeclampsia.

There are currently few options for effective treatment of preeclampsia or interventions for prevention. Study of genetic variants involved in carbon monoxide and nitric oxide signaling can provide additional support for clinical trials that are currently underway investigating potential treatments that operate upon these pathways, such as sildenafil and L-arginine. Pharmacogenomic analysis of data from those clinical trials will likely provide additional insights into mechanisms of effect.

Our study underscores the importance of addressing the contribution of both maternal and fetal components to the condition of preeclampsia. The fetal genotypic contribution to preeclampsia has been widely understudied despite the probable placental contribution to the condition; indeed, our primary finding was for a fetal SNP. Methods that incorporate maternal and fetal genetics may be relevant to a diverse set of reproductive and perinatal conditions that may have a dyadic nature (e.g. congenital anomalies, preterm birth, breastfeeding problems, postpartum depression) most of which have predominantly been explored as solely fetal or solely maternal. Failure to acknowledge both maternal and child genotype may result in missing important associations in which mother and child genotype may operate differently.

This study contributes to the growing area of genetic perinatal epidemiology. Perinatal epidemiology is an area in which there are unique methodological challenges due to unobserved data in early pregnancy, limited ability to collect biomarkers of a fetus, changing maternal behaviors throughout the course of pregnancy, and simultaneous maternal and child contributions. This study brings attention to the importance of considering these issues as genetic epidemiology in the area of perinatal health continues to expand.

APPENDIX: SUPPLEMENTARY TABLES AND FIGURES

Table S1. Genes and pathways for Aim 1

Gene Symbol	Chr	Primary Canonical Pathway
CASP9	1	eNOS Signaling
ADCY3	2	eNOS Signaling
CASP8	2	eNOS Signaling
CASP3	4	eNOS Signaling
CCNA2	4	eNOS Signaling
GUCY1A3	4	eNOS Signaling
GUCY1B3	4	eNOS Signaling
ESR1	6	eNOS Signaling
CAV1	7	eNOS Signaling
ADCY8	8	eNOS Signaling
GUCY1A2	11	eNOS Signaling
CHRNA3	15	eNOS Signaling
CHRNA5	15	eNOS Signaling
EPAS1	2	eNOS Signaling Related Gene
STAT1	2	eNOS Signaling Related Gene
PDE1C	7	eNOS Signaling Related Gene
PDGFA	7	eNOS Signaling Related Gene
PON1	7	eNOS Signaling Related Gene
ENG	9	eNOS Signaling Related Gene
TLR4	9	eNOS Signaling Related Gene
GUCY2C	12	eNOS Signaling Related Gene
MPO	17	eNOS Signaling Related Gene
PRKCA	17	eNOS Signaling Related Gene
PDGFB	22	eNOS Signaling Related Gene
PDGFC	4	eNOS Signaling/HIF1α Signaling
NOS3	7	eNOS Signaling/HIF1α Signaling
PIK3C2G	12	eNOS Signaling/HIF1α Signaling
AKT1	14	eNOS Signaling/HIF1α Signaling
FIGF	X	eNOS Signaling/HIF1α Signaling
MAPK10	4	NOS Signaling/HIF1α Signaling
MAPK9	5	NOS Signaling/HIF1α Signaling
MAPK13	6	NOS Signaling/HIF1α Signaling
MAPK14	6	NOS Signaling/HIF1α Signaling
MAPK8	10	NOS Signaling/HIF1α Signaling
MAPK3	16	NOS Signaling/HIF1α Signaling
MAPK1	22	NOS Signaling/HIF1α Signaling
MAPK11	22	NOS Signaling/HIF1α Signaling
MAPK12	22	NOS Signaling/HIF1α Signaling
NOS1	12	NOS Signaling/HIF1α Signaling
NOS2	17	NOS signaling/HIF1α Signaling
ARNT	1	HIF1α Signaling
NCOA1	2	HIF1α Signaling
EDN1	6	HIF1α Signaling
EPO	7	HIF1α Signaling
MAPK15	8	HIF1α Signaling
MMP1	11	HIF1α Signaling
MMP12	11	HIF1α Signaling
MMP3	11	HIF1α Signaling
MDM2	12	HIF1α Signaling
APEX1	14	HIF1α Signaling
HIF1A	14	HIF1α Signaling
MAPK6	15	HIF1α Signaling
MMP2	16	HIF1α Signaling
MAPK7	17	HIF1α Signaling
MAPK4	18	HIF1α Signaling
EGLN2	19	HIF1α Signaling

Gene Symbol	Chr	Primary Canonical Pathway
MMP9	20	HIF1α Signaling
HBB	11	Heme Degradation
HBA1	16	Heme Degradation
HBA2	16	Heme Degradation
HMOX2	16	Heme Degradation
HMOX1	22	Heme Degradation
CA9	9	Heme Degradation Related Gene
NGB	14	Heme Degradation Related Gene
HP	16	Heme Degradation Related Gene
MB	22	Heme Degradation Related Gene

Table S2. Genes and pathways for Aim 2

Pathway	Description	Genes*
Endothelial nitric oxide synthase signaling	Describes the synthesis of nitric oxide from L-arginine. Endothelial nitric oxide synthase plays a crucial role in the state of blood vessel vasodilation and blood pressure regulation.	*CASP9, ADCY3, CASP8, CASP3, CCNA2, GUCY1A3, GUCY1B3, ESR1, CAV1, ADCY8, GUCY1A2, CHRNA3, CHRNA5, EPAS1, STAT1, PDE1C, PDGFA, PON1, ENG, TLR4, GUCY2C, MPO, PRKCA, PDGFB, PDGFC, NOS3, PIK3C2G, AKT1, FIGF, MAPK10, MAPK9, MAPK13, MAPK14, MAPK8, MAPK3, MAPK1, MAPK11, MAPK12, NOS1, NOS2*
Heme degradation	Describes the breakdown of hemoglobin into carbon monoxide, biliverdin, iron, and bilirubin.	*HBB, HBA1, HBA2, HMOX2, HMOX1, CA9, NGB, HP, MB*
Hypoxia-inducible factor 1-alpha signaling	Describes regulation of oxygen homeostasis and response to hypoxia. Activates transcription of nitric oxide synthase. Also involved in the xenobiotic response via aryl hydrocarbon receptor nuclear translocator.	*PDGFC, NOS3, PIK3C2G, AKT1, FIGF, MAPK10, MAPK9, MAPK13, MAPK14, MAPK8, MAPK3, MAPK1, MAPK11, MAPK12, NOS1, NOS2, ARNT, NCOA1, EDN1, EPO, MAPK15, MMP12, MMP3, MDM2, APEX1, HIF1A, MAPK6, MMP2, MAPK7, MAPK4, EGLN2, MMP9*
Xenobiotic metabolism	Describes the three groups of enzymes that metabolize, eliminate, and detoxify harmful substances. Phase I: introduces polar moiety, Phase II: conjugates toxins to small hydrophilic molecules, Phase III: transporters that export toxins.	*IL1A, MAPK1, HS2ST1, CYP3A7, MAPK13, IL6, NFKB1, MAPK11, PTGES3, ARNT, HMOX1, GSTT1, MAOB, CYP1A2, ALDH1A1, MAPK3, NOS2, AHR, GSTK1, PRKCA, ATM, GSTA3, ABCB1, GSTM1, CYP1A1, GSTM3, GSTA4, NQO1, MAPK8, PIK3C2G, MAPK9, NFKB2, MAPK12, GSTO1, CYP1B1, AIP, MAPK14, SULT1A1, NCOA1, IL1B, CYP2B6, MAPK7, GSTO2, TNF, UGT1A9, GSTP1, MAOA*
Aryl hydrocarbon receptor signaling	Describes mediation of halogenated and polycyclic aromatic hydrocarbons by the aryl hydrocarbon receptor. Activates xenobiotic metabolizing enzymes and other growth factors and proteins involved in cell cycle progression and apoptosis.	*CDKN2A, MAPK1, TP73, IL6, CCND1, ARNT, MYC, RB1, ALDH1A1, CYP1A2, CCND3, TGFB1, MAPK3, AHR, FASLG, GSTK1, ATM, TP53, CCNE2, GSTM3, CDK6, NFKB2, CCND2, E2F1, TGFB3, ESR1, TNF, GSTP1, CDK2, IL1A, CDK4, NFKB1, PTGES3, FAS. CCNA2, GSTT1, TGFB2, CHEK2, GSTA3, GSTM1, CYP1A1, GSTA4, NQO1, APAF1, MAPK8, MDM2, BAX, CYP1B1, GSTO1, AIP, CCNE1, CDKN1A, IL1B, ATR, CDKN1B GSTO2*
Glutathione-mediated detoxification	Describes detoxification in which the first step is catalyzed by glutathione transferases.	*GSTA3, GSTM1, GSTT1, GSTM3, GSTA4, GSTO2, GSTP1, GSTO1, GSTK1*
Nicotine degradation II and III	Describes degradation of nicotine primarily through the metabolic action of cytochrome P450.	*CYP2D6, ADH7, CYP1A1, CYP1A2, CYP2E1, CYP2A7, CYP3A7, CYP2A6, CYP2B6, UGT1A9, CYP1B1*

Table S3. Number of duplicates on each plate

Plate	Duplicate Pairs
NPE003	1
NPE005	1
NPE009	1
NPE010	1
NPE011	1
NPE012	1
NPE013	1
NPE014	1
NPE015	1
NPE016	1
NPE017	1
NPE018	1
NPE019	1
NPE020	1
NPE021	1
NPE022	1
NPE023	1
NPE024	1
NPE025	1
NPE027	1
NPE028	1
NPE029	1
NPE030	1
NPE601	1
NPE602	1
NPE603	1
NPE604	1
NPE605	1
NPE606	1
NPE607	1
NPE608	1
NPE609	1
NPE610	1
NPE611	1
NPE612	1
NPE613	1
NPE614	1
NPE615	1
NPE616	1
NPE618	1
NPE619	2
NPE620	1
NPE621	1
NPE623	1
NPE702	1
NPE703	1
NPE704	1
NPE705	1
617/617r	1

Table S4. Study duplicate pairs

Sample ID	Plate	Well	Duplicate ID	Plate	Well	Relationship
92096	3	H03	91567	3	H08	Child
92098	5	F02	91767	5	H09	Child
91282	9	B01	89205	9	A04	Mother
91283	10	F07	89352	10	E10	Child
89405	11	D05	91284	11	F11	Mother
89493	12	E04	91285	12	B05	Mother
89597	13	A06	91286	13	A01	Mother
91287	14	D02	89660	14	H01	Child
91288	15	G10	89791	15	F06	Mother
91289	16	C06	89878	16	F05	Child
91290	17	C05	90013	17	A11	Mother
91291	18	D10	90100	18	H09	Child
91292	19	H03	90179	19	C08	Mother
91293	20	D01	90222	20	G01	Child
90379	21	G09	91294	21	F02	Mother
91295	22	B09	90498	22	H12	Child
91296	23	F06	90539	23	B06	Mother
90599	24	G01	91297	24	D04	Mother
91298	25	E12	90769	25	D11	Mother
91300	27	F03	90890	27	A03	Child
91301	28	C03	91028	28	E08	Child
91119	29	C08	91302	29	B07	Mother
91303	30	E04	91162	30	F01	Child
96109	601	H03	98671	601	F07	Mother
98672	602	E03	96196	602	H02	Child
98673	603	D06	96358	603	F11	Child
98674	604	C01	96408	604	A06	Child
98675	605	D09	96495	605	A05	Mother
98676	606	F03	96648	606	F12	Child
96653	607	C01	98677	607	B07	Mother
98679	608	F07	96788	608	E06	Child
98680	609	E02	96924	609	G11	Child
98681	610	F08	96938	610	F01	Child
98682	611	D03	97118	611	F12	Child
98683	612	D02	97136	612	B03	Child
98684	613	G03	97257	613	E06	Mother
98685	614	B12	97365	614	B08	Mother
98686	615	F05	97407	615	E01	Mother
97501	616	E01	98687	616	B09	Mother
98689	618	C03	97687	618	C01	Mother
97845	619	E09	98690	619	A06	Mother
98382*	619	H05	102247*	619	B07	Mother
98691	620	D01	97943	620	A10	Mother
98692	621	B10	98022	621	H07	Child
98694	623	H05	98181	623	D04	Mother
102243	702	E09	98282	702	D09	Mother
98437	703	B03	102242	701	C03	Mother
102244	703	F02	98558	703	E02	Mother
102245	704	A03	98346	704	A02	Mother
102246	705	H04	98245	705	G04	Mother
98688	617/617r	E02	97620	617	G04	Child

*Duplicate pair 98382/102247 excluded for discordance

Table S5. Identification of trio members from 5 CEPH Utah families

Family ID	Child ID	Gender	Paternal ID	Maternal ID
1341	6991	Female	6993	6985
1347	10859	Female	11881	11882
1362-A*	10860	Male	11992	11993
1362-B*	10861	Female	11994	11995
1408	10831	Female	12155	12156

CEPH Pedigree information can be found at:
https://catalog.coriell.org/0/sections/collections/nigms/CEPHFamilies.aspx?PgId=49&coll=GM
* 1362-A and 1362-B are two branches of the same pedigree. 1362-A is the father and paternal grandparents while 1362-B is mother and maternal grandparents, therefore, branches are unrelated.

Table S6. Number of Centre d'Etude du Polymorphisme Humain (CEPH) known family trios genotyped on each plate

Plate	Samples
NPE001	2
NPE002	1
NPE003	1
NPE004	1
NPE005	5
NPE006	7
NPE007	3
NPE008	2
NPE009	1
NPE010	1
NPE011	1
NPE012	1
NPE013	1
NPE014	1
NPE015	1
NPE016	1
NPE017	1
NPE018	1
NPE019	1
NPE020	1
NPE021	1
NPE022	1
NPE023	1
NPE024	1
NPE025	1
NPE026	1
NPE027	1
NPE028	1
NPE029	1
NPE030	1
NPE601	1
NPE602	1
NPE603	2
NPE604	2
NPE605	1
NPE606	1
NPE607	2
NPE608	1
NPE609	1
NPE610	1
NPE611	1
NPE612	1
NPE613	1
NPE614	1
NPE615	1
NPE616	1
NPE617	1
NPE618	2
NPE619	3
NPE620	1
NPE621	1
NPE622	1

Plate	Samples
NPE623	2
NPE701	2
NPE702	2
NPE703	2
NPE704	3
NPE705	1
Total	84

Table S7. CEPH Samples and Plates

Sample ID	Study ID	Plate	Well	Pedigree
NA_06985_06B08	06985	6	B08	1341
NA_06985_06D04	06985	6	D04	1341
NA_06985_27B02	06985	27	B02	1341
NA_06985_28G11	06985	28	G11	1341
NA_06985_601C01	06985	601	C01	1341
NA_06985_618D06	06985	618	D06	1341
NA_06985_619F04	06985	619	F04	1341
NA_06991_05D03	06991	5	D03	1341
NA_06991_25B02	06991	25	B02	1341
NA_06991_602F09	06991	602	F09	1341
NA_06991_603H08	06991	603	H08	1341
NA_06991_619C06	06991	619	C06	1341
NA_06993_05D05	06993	5	D05	1341
NA_06993_26G11	06993	26	G11	1341
NA_06993_603E04	06993	603	E04	1341
NA_06993_618C04	06993	618	C04	1341
NA_10831_04G11	10831	4	G11	1408
NA_10831_05B04	10831	5	B04	1408
NA_10831_23B02	10831	23	B02	1408
NA_10831_24G11	10831	24	G11	1408
NA_10831_615C06	10831	615	C06	1408
NA_10831_617A06	10831	617	A06	1408
NA_10831_703F11	10831	703	F11	1408
NA_10859_06G01	10859	6	G01	1347
NA_10859_06G12	10859	6	G12	1347
NA_10859_08G11	10859	8	G11	1347
NA_10859_15B02	10859	15	B02	1347
NA_10859_16G11	10859	16	G11	1347
NA_10859_604B05	10859	604	B05	1347
NA_10859_606B11	10859	606	B11	1347
NA_10859_621D08	10859	621	D08	1347
NA_10859_622E01	10859	622	E01	1347
NA_10860_07E06	10860	7	E06	1362A
NA_10860_09B02	10860	9	B02	1362A
NA_10860_607F01	10860	607	F01	1362A
NA_10860_701A07	10860	701	A07	1362A
NA_10861_02G11	10861	2	G11	1362B
NA_10861_17B02	10861	17	B02	1362B
NA_10861_30G11	10861	30	G11	1362B
NA_10861_610H02	10861	610	H02	1362B
NA_10861_704C06	10861	704	C06	1362B
NA_11881_06D12	11881	6	D12	1347
NA_11881_14G11	11881	14	G11	1347
NA_11881_604H09	11881	604	H09	1347
NA_11881_620D09	11881	620	D09	1347
NA_11882_06G11	11882	6	G11	1347
NA_11882_08G04	11882	8	G04	1347
NA_11882_13B02	11882	13	B02	1347

Sample ID	Study ID	Plate	Well	Pedigree
NA_11882_605C06	11882	605	C06	1347
NA_11882_619H03	11882	619	H03	1347
NA_11992_06H05	11992	6	H05	1362A
NA_11992_11B02	11992	11	B02	1362A
NA_11992_12G11	11992	12	G11	1362A
NA_11992_607C04	11992	607	C04	1362A
NA_11992_609F12	11992	609	F12	1362A
NA_11992_701F09	11992	701	F09	1362A
NA_11993_07B02	11993	7	B02	1362A
NA_11993_07H03	11993	7	H03	1362A
NA_11993_10G11	11993	10	G11	1362A
NA_11993_608E05	11993	608	E05	1362A
NA_11993_702D04	11993	702	D04	1362A
NA_11993_702G05	11993	702	G05	1362A
NA_11994_01F12	11994	1	F12	1362B
NA_11994_03B02	11994	3	B02	1362B
NA_11994_19B02	11994	19	B02	1362B
NA_11994_20G11	11994	20	G11	1362B
NA_11994_29B02	11994	29	B02	1362B
NA_11994_612F01	11994	612	F01	1362B
NA_11994_623D06	11994	623	D06	1362B
NA_11995_01B02	11995	1	B02	1362B
NA_11995_18G11	11995	18	G11	1362B
NA_11995_611A06	11995	611	A06	1362B
NA_11995_613B02	11995	613	B02	1362B
NA_11995_623G03	11995	623	G03	1362B
NA_11995_705B02	11995	705	B02	1362B
NA_12155_05B02	12155	5	B02	1408
NA_12155_21B02	12155	21	B02	1408
NA_12155_614G02	12155	614	G02	1408
NA_12155_703B08	12155	703	B08	1408
NA_12155_704B02	12155	704	B02	1408
NA_12156_05B01	12156	5	B01	1408
NA_12156_22G11	12156	22	G11	1408
NA_12156_616B02	12156	616	B02	1408
NA_12156_704A10	12156	704	A10	1408

Table S8. Discordant CEPH duplicates

		Discordant SNPs	
Sample 1	**Sample 2**	**Chr**	**SNP**
06911 05 D03	06911 619 C06	5	rs7718740
		22	rs4820872
06911 25 B02	06911 619 C06	5	rs7718740
		22	rs4820872
06911 602 F09	06911 619 C06	5	rs7718740
		22	rs4820872

All duplicate problems were due to sample 6991 in plate 619 well C06

Table S9. All sample exclusions

FID	IID	Reason for Exclusion
11169	89205	QC Duplicate
11083	91284	QC Duplicate
10265	91285	QC Duplicate
10137	91286	QC Duplicate
11694	89791	QC Duplicate
12659	90013	QC Duplicate
10982	90179	QC Duplicate
11859	91294	QC Duplicate
10022	90539	QC Duplicate
11396	91297	QC Duplicate
11908	90769	QC Duplicate
10763	91302	QC Duplicate
11269	98671	QC Duplicate
10771	96495	QC Duplicate
11964	98677	QC Duplicate
12361	97257	QC Duplicate
11765	97365	QC Duplicate
11878	97407	QC Duplicate
11757	98687	QC Duplicate
11727	97687	QC Duplicate
12737	98690	QC Duplicate
10143	102247	QC Duplicate
10842	97943	QC Duplicate
10255	98181	QC Duplicate
11501	102242	QC Duplicate
11901	98282	QC Duplicate
10689	98558	QC Duplicate
12621	98346	QC Duplicate
11787	98245	QC Duplicate
10832	96189	Sex Problem - Coded as female genotype male
12072	97549	Sex Problem - Coded as female genotype male
11887	98636	Sex Problem - Coded as female genotype male
11337	90324	Potential sample swap; labeled baby but duplicate of another mom
10471	89191	Fugitive relatedness - related to baby 96774
12293	89193	Fugitive relatedness - related to mom 89745
12245	96349	Fugitive relatedness - related to mom 89225
12005	97516	Fugitive relatedness - related to mom 89203
10553	89340	Fugitive relatedness - related to mom 90631
10296	89359	Fugitive relatedness - related to mom 96871 and baby 96872
10296	89360	Fugitive relatedness - related to mom 96871 and baby 96872
10403	98231	Fugitive relatedness - related to mom 89329
10115	89369	Fugitive relatedness - related to mom 96557 and baby 96558
10115	89370	Fugitive relatedness - related to mom 96557
10488	96605	Fugitive relatedness - related to mom 89433
12723	89489	Fugitive relatedness - related to baby 96872
11246	91133	Fugitive relatedness - related to mom 89549 and baby 89550
11246	91134	Fugitive relatedness - related to mom 89549
11782	97593	Fugitive relatedness - related to mom 89635 and baby 89636
11782	97594	Fugitive relatedness - related to mom 89635
11481	89797	Fugitive relatedness - related to mom 97753 and baby 97754
11481	89798	Fugitive relatedness - related to mom 97753
11519	89799	Fugitive relatedness - related to baby 97554

FID	IID	Reason for Exclusion
11646	98410	Fugitive relatedness - related to mom 90021
10627	90009	Fugitive relatedness - related to baby 97690
10441	89971	Fugitive relatedness - related to mom 90717 and baby 90718
10441	89972	Fugitive relatedness - related to mom 90717 and baby 90718
12167	98631	Fugitive relatedness - related to mom 90219 and baby 90220
11892	97454	Fugitive relatedness - related to baby 91293 and mom 90221
12203	90284	Fugitive relatedness - related to baby 90650
12521	97277	Fugitive relatedness - related to mom 90505
10760	98224	Fugitive relatedness - related to mom 90741
11864	97659	Fugitive relatedness - related to baby 90814
11864	97660	Fugitive relatedness - related to baby 90814
12037	91165	Fugitive relatedness - related to mom 91229
10490	91236	Fugitive relatedness - related to mom 97723
10490	91235	Fugitive relatedness - related to mom 97723 and baby 97724
12685	98363	Fugitive relatedness - related to mom 91201
10225	96783	Fugitive relatedness - related to baby 96108
10998	96242	Fugitive relatedness - related to mom 97167
11781	97461	Fugitive relatedness - related to mom 96639
11845	97031	Fugitive relatedness - related to baby 96916
11845	97032	Fugitive relatedness - related to baby 96916
12062	98653	Fugitive relatedness - related to baby 97426
10355	97918	Fugitive relatedness - related to baby 97912
11083	89405	Shown as unrelated for FID 11083
11083	89406	Shown as unrelated for FID 11083
11723	89437	Shown as unrelated for FID 11723
11723	89438	Shown as unrelated for FID 11723
11528	89651	Shown as unrelated for FID 11528
11528	89652	Shown as unrelated for FID 11528
12309	90335	Shown as unrelated for FID 12309
12309	90336	Shown as unrelated for FID 12309
10344	91299	Shown as unrelated for FID 10344
10344	90809	Shown as unrelated for FID 10344
10837	93526	Shown as unrelated for FID 10837
10837	93527	Shown as unrelated for FID 10837
12037	91166	Shown as unrelated for FID 12037
11822	96757	Shown as unrelated for FID 11822
11822	96758	Shown as unrelated for FID 11822
10143	98382	Duplicate with high levels of discordance (duplicate of 10143, 102247)
11757	97501	Duplicate with high levels of discordance (duplicate of 11757, 98687)

Table S10. SNPs missing genotype information for all individuals

SNP	Chr	Freq Missing
UNC_Eclampsia_chr1:11745321	1	1
UNC_Eclampsia_chr1:11860798	1	1
exm17102	1	1
exm19516	1	1
exm2253585	1	1
exm75804	1	1
exm81979	1	1
exm90310	1	1
exm90327	1	1
exm101774	1	1
exm136342	1	1
rs2689154	1	1
indel.11236	1	1
newrs11339452	1	1
rs17015608	2	1
exm243367	2	1
rs892515	2	1
UNC_Eclampsia_chr3:12256739	3	1
exm332686	3	1
exm2238321	3	1
exm342498	3	1
UNC_Eclampsia_rs1840961	3	1
UNC_Eclampsia_rs1823227	3	1
exm371798	3	1
exm384902	4	1
exm402024	4	1
UNC_Eclampsia_rs4693753	4	1
UNC_Eclampsia_rs2250724	4	1
exm427425	4	1
exm476743	5	1
exm522796	6	1
UNC_Eclampsia_rs404655	6	1
UNC_Eclampsia_rs2735096	6	1
rs2571391	6	1
exm-rs3129813	6	1
exm528518	6	1
exm-rs3130970	6	1
exm-rs1265095	6	1
UNC_Eclampsia_rs2770	6	1
exm-rs10046127	6	1
exm535710	6	1
exm535831	6	1
UNC_Eclampsia_rs34951355	6	1
exm536131	6	1
exm-rs2144014	6	1
exm-rs6900824	6	1
exm540227	6	1
exm546933	6	1
exm548257	6	1
exm555771	6	1
exm561835	6	1
exm588851	6	1
exm612435	7	1
UNC_Eclampsia_chr7:71788534	7	1
exm636185	7	1
exm637995	7	1
exm644742	7	1
exm658486	7	1
exm2258137	7	1
exm684450	8	1

SNP	Chr	Freq Missing
exm695749	8	1
UNC_Eclampsia_rs16891561	8	1
exm700589	8	1
rs12265	8	1
exm713312	8	1
exm730833	8	1
exm742955	9	1
rs2383246	9	1
exm790118	9	1
exm790686	9	1
UNC_Eclampsia_chr9:135034420	9	1
exm820363	10	1
UNC_Eclampsia_rs250706	10	1
exm854205	10	1
UNC_Eclampsia_rs1706879	11	1
exm2219045	11	1
exm966579	11	1
var_11_125864230	11	1
exm982429	12	1
rs4763494	12	1
indel.24166	12	1
exm1018827	12	1
rs373767	13	1
rs577487	13	1
exm1085830	14	1
exm2251753	14	1
variant.31710	14	1
UNC_Eclampsia_rs181478	14	1
exm1113017	14	1
exm1129029	14	1
exm1134125	14	1
exm1166050	15	1
exm1178847	15	1
exm1325307	17	1
exm1912211	18	1
exm1420231	19	1
exm1422205	19	1
UNC_Eclampsia_chr19:19402787	19	1
indel.54461	19	1
var_19_22272062	19	1
UNC_Eclampsia_rs12461383	19	1
variant.56280	19	1
exm1477565	19	1
exm1478264	19	1
exm1504396	19	1
exm1504410	19	1
exm1506322	19	1
exm-IND19-60217915	19	1
newrs143207252	19	1
exm1518874	20	1
exm1536190	20	1
UNC_Eclampsia_rs495064	22	1
exm1632722	23	1
indel.108572	23	1
exm1653205	23	1
newrs237520	23	1
UNC_Eclampsia_mt3796	26	1
exm2263340	26	1
200610-25	26	1
2010-08-MT-158	26	1

SNP	Chr	Freq Missing
MitoT13966C	26	1
exm2263330	26	1
UNC_Eclampsia_mt14769	26	1
exm2216494	26	0.09315

Table S11. Sensitivity analysis results.

Table S11a. Sensitivity analysis results. Summary of SNPs presented in paper 1 excluding samples with first principal component >0.04 (n=1,994 mother-child dyads)

						Mother			Child			Joint Test	
Marker	**Chr**	**Position**	**MAF**	**Gene**	**Genotype**	**RR (95% CI)**	**p value**	**FDR q value**	**RR (95% CI)**	**p value**	**FDR q value**	**p value**	**FDR q value**
rs722208	6	152322885	0.30	*ESR1*	*GG*	Referent	4.30×10^{-5}	0.03	Referent	0.58	0.87	4.44×10^{-4}	0.09
					AG	0.64 (0.52, 0.79)			1.06 (0.89, 1.27)				
					AA	0.93 (0.66, 1.33)			1.20 (0.85, 1.71)				
rs1569788	6	152328616	0.30	*ESR1*	*TT*	Referent	2.16×10^{-5}	0.03	Referent	0.60	0.87	2.34×10^{-4}	0.07
					CT	0.64 (0.52, 0.78)			1.05 (0.88, 1.27)				
					CC	0.95 (0.67, 1.35)			1.20 (0.84, 1.70)				
rs3020365	6	152367993	0.36	*ESR1*	*GG*	Referent	7.09×10^{-5}	0.03	Referent	0.33	0.87	6.53×10^{-4}	0.10
					TG	0.65 (0.53, 0.79)			1.13 (0.94, 1.35)				
					TT	0.88 (0.64, 1.21)			1.24 (0.90, 1.70)				
rs3020366	6	152368758	0.37	*ESR1*	*TT*	Referent	7.73×10^{-5}	0.03	Referent	0.46	0.87	7.92×10^{-4}	0.11
					CT	0.65 (0.53, 0.80)			1.09 (0.91, 1.31)				
					CC	0.88 (0.64, 1.20)			1.21 (0.89, 1.66)				
rs6470860	8	131905190	0.41	*ADCY8*	*AA*	Referent	0.84	0.94	Referent	1.17×10^{-4}	0.12	5.02×10^{-5}	0.03
					GA	1.06 (0.86, 1.31)			1.36 (1.13, 1.63)				
					GG	1.03 (0.76, 1.40)			1.89 (1.39, 2.57)				
rs12547243	8	131921956	0.29	*ADCY8*	*GG*	Referent	0.03	0.64	Referent	1.08×10^{-5}	0.01	1.19×10^{-7}	1.44×10^{-4}
					AG	1.19 (0.97, 1.45)			1.45 (1.22, 1.72)				
					AA	0.78 (0.53, 1.13)			2.16 (1.49, 3.14)				
rs7459573	8	131928401	0.34	*ADCY8*	*AA*	Referent	0.84	0.94	Referent	2.91×10^{-4}	0.19	2.91×10^{-4}	0.07
					GA	1.06 (0.87, 1.30)			1.41 (1.18, 1.68)				
					GG	1.02 (0.73, 1.43)			1.61 (1.17, 2.23)				

Table S11b. Sensitivity analysis results. Summary of SNPs presented in paper 1 excluding samples with first principal component >0.01 =1,964 mother-child dyads)

						Mother			Child			Joint Test	
Marker	**Chr**	**Position**	**MAF**	**Gene**	**Genotype**	**RR (95% CI)**	**p value**	**FDR q value**	**RR (95% CI)**	**p value**	**FDR q value**	**p value**	**FDR q value**
rs722208	6	152322885	0.30	*ESR1*	*GG*	Referent	6.09×10^{-5}	0.02	Referent	0.60	0.85	5.68×10^{-4}	0.08
					AG	0.64 (0.52, 0.79)			1.05 (0.87, 1.26)				
					AA	0.91 (0.64, 1.30)			1.20 (0.84, 1.71)				
rs1569788	6	152328616	0.30	*ESR1*	*TT*	Referent	3.12×10^{-5}	0.02	Referent	0.62	0.85	3.05×10^{-4}	0.08
					CT	0.64 (0.52, 0.78)			1.04 (0.87, 1.25)				
					CC	0.93 (0.65, 1.32)			1.19 (0.84, 1.70)				
rs3020365	6	152367993	0.36	*ESR1*	*GG*	Referent	5.77×10^{-5}	0.02	Referent	0.43	0.84	5.90×10^{-4}	0.08
					TG	0.64 (0.52, 0.79)			1.11 (0.93, 1.33)				
					TT	0.86 (0.63, 1.18)			1.21 (0.88, 1.66)				
rs3020366	6	152368758	0.37	*ESR1*	*TT*	Referent	6.20×10^{-5}	0.02	Referent	0.57	0.85	6.62×10^{-4}	0.08
					CT	0.64 (0.52, 0.79)			1.07 (0.89, 1.29)				
					CC	0.86 (0.63, 1.18)			1.18 (0.86, 1.62)				
rs6470860	8	131905190	0.41	*ADCY8*	*AA*	Referent	0.82	0.93	Referent	1.78×10^{-4}	0.12	6.75×10^{-5}	0.04
					GA	1.06 (0.86, 1.31)			1.38 (1.15, 1.65)				
					GG	1.01 (0.74, 1.37)			1.88 (1.38, 2.57)				
rs12547243	8	131921956	0.29	*ADCY8*	*GG*	Referent	0.02	0.58	Referent	9.6×10^{-6}	0.01	6.79×10^{-8}	7.64×10^{-5}
					AG	1.20 (0.98, 1.47)			1.46 (1.23, 1.73)				
					AA	0.76 (0.52, 1.10)			2.18 (1.49, 3.19)				
rs7459573	8	131928401	0.34	*ADCY8*	*AA*	Referent	0.83	0.93	Referent	6.67×10^{-4}	0.22	5.16×10^{-4}	0.08
					GA	1.06 (0.87, 1.30)			1.40 (1.17, 1.67)				
					GG	1.01 (0.73, 1.42)			1.59 (1.15, 2.20)				

Table S12. Replication Results

Table S12a. Replication results for replication cohort of preeclampsia

				Mother		Child	
Marker	**Chr**	**Gene**	**Association**	**RR (95% CI)**	**p value**	**RR (95% CI)**	**p value**
PE Associations (Overall)							
rs7435347	4	*GUCY1A3*	PE (Mother)	0.96 (0.86, 1.06)	0.44	0.98 (0.86, 1.11)	0.73
rs1569788	6	*ESR1*	PE (Mother)	0.98 (0.90, 1.07)	0.62	0.96 (0.86, 1.08)	0.52
rs3020366	6	*ESR1*	PE (Mother)	0.97 (0.89, 1.06)	0.50	0.98 (0.88, 1.08)	0.66
rs6462324	7	*PDE1C*	PE (Mother)	0.93 (0.86, 1.01)	0.07	0.91 (0.82, 1.01)	0.07
rs6470860	8	*ADCY8*	PE (Child)	1.00 (0.93, 1.09)	0.07	0.94 (0.86, 1.04)	0.26
rs12547243	8	*ADCY8*	PE (Child)	0.96 (0.89, 1.05)	0.39	0.97 (0.87, 1.07)	0.54
rs7459573	8	*ADCY8*	PE (Child)	0.95 (0.88, 1.03)	0.23	0.93 (0.84, 1.03)	0.14
rs17475920	12	*PIK3C2G*	PE (Mother)	1.03 (0.91, 1.18)	0.60	0.93 (0.79, 1.09)	0.39
rs9634063	12	*PIK3C2G*	PE (Mother)	1.03 (0.92, 1.16)	0.57	0.93 (0.80, 1.08)	0.34
Severe Preeclampsia Associations							
No SNPs with $Q \leq 0.2$							
Early-onset Preeclampsia Associations							
No SNPs with $Q \leq 0.2$							
Preeclampsia with Early Delivery Associations							
rs30593	7	*PDE1C*	PE + early delivery (Mother)	1.00 (0.92, 1.08)	0.98	0.90 (0.82, 1.00)	0.05
rs12785615	11	*GUCY1A2*	PE + early delivery (Mother)	1.06 (0.95, 1.18)	0.32	0.93 (0.81, 1.07)	0.30
rs1455590	11	*GUCY1A2*	PE + early delivery (Mother)	1.03 (0.94, 1.13)	0.51	0.93 (0.82, 1.04)	0.19
rs11636443	15	*MAPK6*	PE + early delivery (Mother)	0.97 (0.90, 1.05)	0.48	1.05 (0.95, 1.16)	0.32
Preeclampsia Accompanied by Small for Gestational Age Associations							
rs30562	7	*PDE1C*	PE + SGA (Child)	1.00 (0.92, 1.08)	0.96	0.94 (0.85, 1.04)	0.20

Table S12b. Replication results for replication cohort of early preeclampsia

				Mother		Child	
Marker	**Chr**	**Gene**	**Association**	**RR (95% CI)**	**p value**	**RR (95% CI)**	**p value**
PE Associations (Overall)							
rs7435347	4	*GUCY1A3*	PE (Mother)	1.14 (0.96, 1.36)	0.15	1.07 (0.86, 1.35)	0.54
rs1569788	6	*ESR1*	PE (Mother)	1.07 (0.92, 1.25)	0.37	1.01 (0.83, 1.23)	0.92
rs3020366	6	*ESR1*	PE (Mother)	1.01 (0.87, 1.16)	0.94	0.96 (0.79, 1.15)	0.62
rs6462324	7	*PDE1C*	PE (Mother)	0.97 (0.84, 1.11)	0.64	1.00 (0.83, 1.21)	0.99
rs6470860	8	*ADCY8*	PE (Child)	1.00 (0.87, 1.15)	0.99	0.94 (0.78, 1.12)	0.46
rs12547243	8	*ADCY8*	PE (Child)	0.94 (0.81, 1.09)	0.40	0.87 (0.72, 1.06)	0.16
rs7459573	8	*ADCY8*	PE (Child)	0.95 (0.82, 1.09)	0.45	0.88 (0.73, 1.06)	0.17
rs17475920	12	*PIK3C2G*	PE (Mother)	1.17 (0.94, 1.45)	0.17	0.93 (0.70, 1.25)	0.64
rs9634063	12	*PIK3C2G*	PE (Mother)	1.15 (0.95, 1.41)	0.17	0.95 (0.73, 1.25)	0.72
Severe Preeclampsia Associations							
No SNPs with $Q \leq 0.2$							
Early-onset Preeclampsia Associations							
No SNPs with $Q \leq 0.2$							
Preeclampsia with Early Delivery Associations							
rs30593	7	*PDE1C*	PE + early delivery (Mother)	1.00 (0.87, 1.16)	0.99	0.87 (0.72, 1.05)	0.14
rs12785615	11	*GUCY1A2*	PE + early delivery (Mother)	0.99 (0.81, 1.20)	0.88	0.92 (0.71, 1.19)	0.51
rs1455590	11	*GUCY1A2*	PE + early delivery (Mother)	1.00 (0.85, 1.18)	0.95	0.90 (0.73, 1.11)	0.32
rs11636443	15	*MAPK6*	PE + early delivery (Mother)	0.98 (0.85, 112)	0.72	1.02 (0.85, 1.21)	0.86
Preeclampsia Accompanied by Small for Gestational Age Associations							
rs30562	7	*PDE1C*	PE + SGA (Child)	1.00 (0.86, 1.15)	0.95	0.95 (0.79, 1.15)	0.62

Table S13.

Table S13a. Relative risks for the association of smoking and preeclampsia for maternal and child smoking-SNP interactions stratified by genotype, in the sensitivity analysis model allowing smoking exposure to vary by mating type. Results are presented for SNPs with interaction $p < 0.001$ in the original model.

						Mother		Child	
Marker[a]	**Chr**	**Position**	**MAF**	**Gene**	**Genotype**	**RR (95% CI)**	**P-Interaction**	**RR (95% CI)**	**P-Interaction**
rs3765692	1	3584771	0.22	*TP73*	*TT*	1.16 (0.81, 1.66)	5.5×10^{-4}	1.02 (0.71, 1.46)	0.03
					CT	0.45 (0.29, 0.71)		0.58 (0.39, 0.88)	
					CC	0.18 (0.11, 0.30)		0.34 (0.21, 0.53)	
rs10770343	12	18414253	0.31	*PIK3C2G*	*AA*	0.99 (0.71, 1.39)	0.07	0.99 (0.70, 1.42)	0.12
					CA	0.67 (0.43, 1.05)		0.69 (0.45, 1.06)	
					CC	0.46 (0.27, 0.78)		0.48 (0.30, 0.79)	
rs2278361	12	99043207	0.21	*APAF1*	*TT*	0.63 (0.44, 0.91)	0.04	0.68 (0.47, 0.99)	0.20
					CT	1.01 (0.69, 1.47)		0.93 (0.64, 1.37)	
					CC	1.62 (1.10, 2.37)		1.29 (0.88, 1.89)	

Table S13b. Relative risks for the association of smoking and preeclampsia for all maternal and child smoking-SNP interactions stratified by genotype, in the model allowing smoking exposure to vary by mating type. Results are presented for SNPs with interaction $p < 0.001$ in the sensitivity analysis model allowing smoking exposure to vary by mating type.

						Mother		Child	
Marker[a]	**Chr**	**Position**	**MAF**	**Gene**	**Genotype**	**RR (95% CI)**	**P-Interaction**	**RR (95% CI)**	**P-Interaction**
rs3765692	1	3584771	0.22	*TP73*	*TT*	1.16 (0.81, 1.66)		1.02 (0.71, 1.46)	
					CT	0.45 (0.29, 0.71)	5.5×10^{-4}	0.58 (0.39, 0.88)	0.03
					CC	0.18 (0.11, 0.30)		0.34 (0.21, 0.53)	
rs995647	2	24810255	0.13	*NCOA1*	*AA*	0.66 (0.48, 0.91)		0.58 (0.42, 0.81)	
					GA	1.54 (0.85, 2.80)	0.10	1.93 (1.14, 3.27)	3.0×10^{-5}
					GG	3.61 (1.65, 7.88)		6.39 (3.27, 12.47)	
rs7929753	11	106795127	0.25	*GUCY1A2*	*CC*	1.00 (0.68, 1.46)		1.38 (0.94, 2.03)	
					TC	0.68 (0.48, 0.97)	0.01	0.50 (0.34, 0.73)	1.2×10^{-4}
					TT	0.47 (0.34, 0.64)		0.18 (0.12, 0.26)	

Table S14. Relative risks for the association of smoking and preeclampsia for all maternal and child smoking-SNP interactions with p-interaction < 0.001 in the original analysis (of all smokers), stratified by genotype. Smokers were included as: 1) only smokers who quit smoking during 11-20 weeks' gestation, 2) only smokers who did not quit during 11-20 weeks' gestation, 3) any smokers during 11-20 weeks' gestation.

					Quit Smoking (n=63 smokers)		Did Not Quit (n=151 smokers)		All Smokers (n=214 smokers)	
Marker[a]	**Chr**	**MAF**	**Gene**	**Alleles[1]**	**RR (95% CI)[2]**	**P-Int**	**RR (95% CI)[2]**	**P-Int**	**RR (95% CI)[2]**	**P-Int**
rs3765692	1	0.22	*TP73*	*T/C*	0.62 (0.33, 1.16)	0.12	0.52 (0.32, 0.85)	5.7×10^{-3}	0.53 (0.36, 0.78)	5.9×10^{-4}
rs10770343	12	0.31	*PIK3C2G*	*A/C*	0.66 (0.39, 1.11)	0.10	0.52 (0.33, 0.81)	2.8×10^{-3}	0.59 (0.43, 0.81)	5.9×10^{-4}
rs2278361	12	0.21	*APAF1*	*T/C*	1.31 (0.77, 2.20)	0.33	1.97 (1.38, 2.81)	2.6×10^{-4}	1.68 (1.25, 2.26)	7.5×10^{-4}

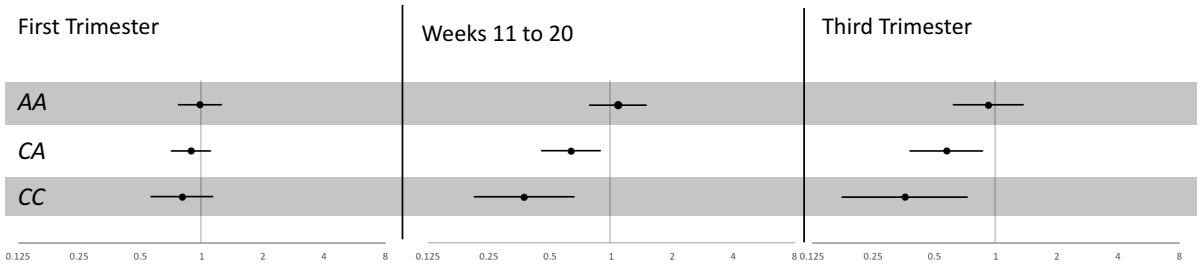
[1]Major allele/minor allele

[2]RR per minor allele

Table S15. Data table for rs3765692 (*TP73*)

Phenotype	Smoking	Mom Alleles	Child Alleles	Count
Control	No	0	0	503
Control	No	0	1	75
Control	No	1	0	99
Control	No	1	1	98
Control	No	1	2	13
Control	No	2	1	11
Control	No	2	2	1
Case	No	0	0	575
Case	No	0	1	81
Case	No	1	0	103
Case	No	1	1	125
Case	No	1	2	15
Case	No	2	1	17
Case	No	2	2	3
Control	Yes	0	0	47
Control	Yes	0	1	12
Control	Yes	1	0	11
Control	Yes	1	1	13
Control	Yes	1	2	1
Control	Yes	2	1	1
Control	Yes	2	2	0
Case	Yes	0	0	62
Case	Yes	0	1	7
Case	Yes	1	0	7
Case	Yes	1	1	6
Case	Yes	1	2	2
Case	Yes	2	1	0
Case	Yes	2	2	0
Control	No	0	Unknown	23
Control	No	1	Unknown	9
Control	No	2	Unknown	2
Case	No	0	Unknown	292
Case	No	1	Unknown	92
Case	No	2	Unknown	10
Control	Yes	0	Unknown	6
Control	Yes	1	Unknown	3
Control	Yes	2	Unknown	0
Case	Yes	0	Unknown	31
Case	Yes	1	Unknown	4
Case	Yes	2	Unknown	0
Control	No	Unknown	0	14
Control	No	Unknown	1	6
Control	No	Unknown	2	0
Case	No	Unknown	0	13
Case	No	Unknown	1	5
Case	No	Unknown	2	1
Control	Yes	Unknown	0	0
Control	Yes	Unknown	1	1
Control	Yes	Unknown	2	0
Case	Yes	Unknown	0	0
Case	Yes	Unknown	1	0
Case	Yes	Unknown	2	0

Figure S1. Risk ratios of the association of preeclampsia in smokers compared to non-smokers, stratified by maternal genotype for rs1077343 (*PIK3C2G*) at three time points: 1) first trimester, 2) gestational weeks 11 to 20 (primary analysis), and 3) third trimester.



REFERENCES

1. World Health Organization. *WHO Recommendations for Prevention and Treatment of Pre-Eclampsia and Eclampsia*. Geneva; 2011.

2. Sibai B, Dekker G, Kupferminc M. Pre-eclampsia. *The Lancet*. 2005;365:785-799.

3. Coomarasamy A, Honest H, Papaioannou S, Gee H, Khan KS. Aspirin for prevention of preeclampsia in women with historical risk factors: A systematic review. *Obstetrics and Gynecology*. 2003;101(6):1319-1332.

4. Eiland E, Nzerue C, Faulkner M. Preeclampsia 2012. *Journal of pregnancy*. 2012.

5. Redman CW, Sargent IL. Latest advances in understanding preeclampsia. *Science (New York, NY)*. 2005;308(5728):1592-1594.

6. England L, Zhang J. Smoking and risk of preeclampsia: a systematic review. *Frontiers in Bioscience*. 2007;12:2471-2483.

7. Ryter SW, Otterbein LE, Morse D, Choi AMK. Heme oxygenase/carbon monoxide signaling pathways: regulation and functional significance. *Molecular and cellular biochemistry*. 2002;234-235(1-2):249-263.

8. Bainbridge S a., Smith GN. HO in pregnancy. *Free Radical Biology and Medicine*. 2005;38(8):979-988.

9. Bacsik Z, McGregor J, Mink J. FTIR analysis of gaseous compounds in the mainstream smoke of regular and light cigarettes. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*. 2007;45(2):266-271.

10. Parrish ME, Lyons-Hart JL, Shafer KH. Puff-by-puff and intrapuff analysis of cigarette smoke using infrared spectroscopy. *Vibrational Spectroscopy*. 2001;27(1):29-42.

11. Wu L, Wang R. Carbon monoxide: endogenous production, physiological functions, and pharmacological applications. *Pharmacological reviews*. 2005;57(4):585-630.

12. Ndisang JF, Tabien HEN, Wang R. Carbon monoxide and hypertension. *Journal of hypertension*. 2004;22(6):1057-1074.

13. Siow RCM, Sato H, Mann GE. Heme oxygenase-carbon monoxide signalling pathway in atherosclerosis: Anti-atherogenic actions of bilirubin and carbon monoxide? *Cardiovascular Research*. 1999;41(2):385-394.

14. Morse D, Sethi J. Carbon monoxide and human disease. *Antioxidants & redox signaling*. 2002;4(2):331-338.

15. Krause BJ, Hanson MA, Casanello P. Role of nitric oxide in placental vascular development and function. *Placenta*. 2011;32(11):797-805.

16. Choi JW, Im MW, Pai SH. Nitric oxide production increases during normal pregnancy and

decreases in preeclampsia. *Annals of clinical and laboratory science*. 2002;32(3):257-263.

17. Casas JP, Cavalleri GL, Bautista LE, Smeeth L, Humphries SE, Hingorani a. D. Endothelial Nitric Oxide Synthase Gene Polymorphisms and Cardiovascular Disease: A HuGE Review. *American Journal of Epidemiology*. 2006;164(10):921-935.

18. Motterlini R, Otterbein LE. The therapeutic potential of carbon monoxide. *Nature reviews Drug discovery*. 2010;9(9):728-743.

19. Ono K, Mannami T, Iwai N. Association of a promoter variant of the haeme oxygenase-1 gene with hypertension in women. *Journal of hypertension*. 2003;21(8):1497-1503.

20. Chen Y, Lin S, Lin M, Tsai H, Kuo S. Microsatellite polymorphism in promoter of heme oxygenase-1 gene is associated with susceptibility to coronary artery disease in type 2 diabetic patients. *Human genetics*. 2002:1-8.

21. Yang L, Quan S, Nasjletti A, Laniado-Schwartzman M, Abraham NG. Heme oxygenase-1 gene expression modulates angiotensin II-induced increase in blood pressure. *Hypertension*. 2004;43(6):1221-1226.

22. Choi SK, Hwang JY, Lee J, et al. Gene expression of heme oxygenase-1 and nitric oxide synthase on trophoblast of preeclampsia. *Korean Journal of Obstetrics and Gynecology*. 2011;54(7):341.

23. Ozturk E, Balat O, Pehlivan S, et al. Endothelial nitric oxide synthase gene polymorphisms in preeclampsia with or without eclampsia in a Turkish population. *Journal of Obstetrics and Gynaecology Research*. 2011;37(12):1778-1783.

24. Enquobahrie D, Meller M, Rice K, Psaty BM, Siscovick DS, Williams M a. Differential placental gene expression in preeclampsia. *American Journal of Obstetrics and Gynecology*. 2008;199(5):566.e1-11.

25. ACOG. Executive Summary: Hypertension in Pregnancy. *Obstetrics and gynecology*. 2013;122(5):1122-1131.

26. Bell MJ. A Hisotorical Overview of Preeclampsia-Eclampsia. *J Obstet Gynecol Neonatal Nurs*. 2010;39(5):510-518.

27. American College of Obstetricians and Gynecologists. ACOG practice bulletin. Diagnosis and management of preeclampsia and eclampsia. Number 33, January 2002. American College of Obstetricians and Gynecologists. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics*. 2002;77(1):67-75.

28. Zhang J, Merialdi M, Platt LD, Kramer MS. Defining normal and abnormal fetal growth: promises and challenges. *American Journal of Obstetrics and Gynecology*. 2010;202(6):522-528.

29. Wallis AB, Saftlas AF, Hsia J, Atrash HK. Secular trends in the rates of preeclampsia, eclampsia, and gestational hypertension, United States, 1987-2004. *American journal of*

hypertension. 2008;21(5):521-526.

30. Dahlstrøm BL, Engh ME, Bukholm G, Oian P. Changes in the prevalence of pre-eclampsia in Akershus County and the rest of Norway during the past 35 years. *Acta obstetricia et gynecologica Scandinavica*. 2006;85(8):916-921.

31. Klungsøyr K, Morken NH, Irgens L, Vollset SE, Skjærven R. Secular trends in the epidemiology of pre-eclampsia throughout 40 years in Norway: Prevalence, risk factors and perinatal survival. *Paediatric and Perinatal Epidemiology*. 2012;26(3):190-198.

32. Heard AR, Dekker G, Chan A, Jacobs DJ, Vreeburg S a, Priest KR. Hypertension during pregnancy in South Australia, part 1: pregnancy outcomes. *The Australian & New Zealand journal of obstetrics & gynaecology*. 2004;44(5):404-409.

33. Hauth JC, Ewell MG, Levine RJ, et al. Pregnancy outcomes in healthy nulliparas who developed hypertension. *Obstetrics and Gynecology*. 2000;95(1):24-28.

34. Lenfant C. Working group report on high blood pressure in pregnancy. *Journal of clinical hypertension (Greenwich, Conn)*. 2001;3(2):75-88.

35. Khan KS, Wojdyla D, Say L, Gülmezoglu a M, Van Look PF. WHO analysis of causes of maternal death: a systematic review. *Lancet*. 2006;367(9516):1066-1074.

36. Duley L. The Global Impact of Pre-eclampsia and Eclampsia. *Seminars in Perinatology*. 2009;33(3):130-137.

37. Barton JR, Sibai BM. Prediction and prevention of recurrent preeclampsia. *Obstetrics and gynecology*. 2008;112(2 Pt 1):359-372.

38. Bramham K, Briley AL, Seed P, Poston L, Shennan AH, Chappell LC. Adverse maternal and perinatal outcomes in women with previous preeclampsia: A prospective study. *American Journal of Obstetrics and Gynecology*. 2011;204(6):512.e1-9.

39. Lykke J, Langhoff-Roos J, Sibai BM, Funai EF, Triche EW, Paidas MJ. Hypertensive pregnancy disorders and subsequent cardiovascular morbidity and type 2 diabetes mellitus in the mother. *Hypertension*. 2009;53(6):944-951.

40. Bellamy L, Casas J-P, Hingorani AD, Williams DJ. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. *BMJ (Clinical research ed)*. 2007;335(7627):974.

41. McDonald SD, Malinowski A, Zhou Q, Yusuf S, Devereaux PJ. Cardiovascular sequelae of preeclampsia/eclampsia: A systematic review and meta-analyses. *American Heart Journal*. 2008;156(5):918-930.

42. Hutcheon J, Lisonkova S, Joseph KS. Epidemiology of pre-eclampsia and the other hypertensive disorders of pregnancy. *Best Practice and Research: Clinical Obstetrics and Gynaecology*. 2011;25(4):391-403.

43. Norwitz E, Repke J. Preeclampsia: Management and Prognosis. *UpToDate*. 2012.

44. Meis PJ, Goldenberg RL, Mercer BM, et al. The preterm prediction study: risk factors for indicated preterm births. Maternal-Fetal Medicine Units Network of the National Institute of Child Health and Human Development. *American journal of obstetrics and gynecology*. 1998;178(3):562-567.

45. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet*. 2008;371(9608):261-269.

46. Sibai BM. Prevention of preeclampsia: a big disappointment. *American journal of obstetrics and gynecology*. 1998;179(5):1275-1278.

47. Dekker G, Sibai B. Primary, secondary, and tertiary prevention of pre-eclampsia. *Lancet*. 2001;357(9251):209-215.

48. Conde-Agudelo A, Villar J, Lindheimer M. World Health Organization systematic review of screening tests for preeclampsia. *Obstetrics and gynecology*. 2004;104(6):1367-1391.

49. Chappell LC, Seed PT, Briley AL, et al. Effect of antioxidants on the occurrence of pre-eclampsia in women at increased risk: A randomised trial. *Lancet*. 1999;354(9181):810-816.

50. Mattar F, Sibai BM. Prevention of preeclampsia. *Seminars in perinatology*. 1999;23(1):58-64.

51. Duley L. Do women with pre-eclampsia, and their babies, benefit from magnesium sulphate? The Magpie Trial: A randomised placebo-controlled trial. *Lancet*. 2002;359(9321):1877-1890.

52. Livingston JC, Livingston LW, Ramsey R, Mabie BC, Sibai BM. Magnesium sulfate in women with mild preeclampsia: A randomized controlled trial. *Obstetrics and Gynecology*. 2003;101(2):217-220.

53. Lucas MJ, Leveno KJ, Cunningham FG. *A Comparison of Magnesium Sulfate with Phenytoin for the Prevention of Eclampsia.*; 1995.

54. Sibai BM. Magnesium sulfate prophylaxis in preeclampsia: Lessons learned from recent trials. *American Journal of Obstetrics and Gynecology*. 2004;190(6):1520-1526.

55. Roberts JM, Taylor RN, Musci TJ, Rodgers GM, Hubel CA, McLaughlin MK. Preeclampsia: an endothelial cell disorder. *American journal of obstetrics and gynecology*. 1989;161(5):1200-1204.

56. Gilbert JS, Ryan MJ, LaMarca BB, Sedeek M, Murphy SR, Granger JP. Pathophysiology of hypertension during preeclampsia: linking placental ischemia with endothelial dysfunction. *American journal of physiology Heart and circulatory physiology*. 2008;294(2):H541-H550.

57. Powe CE, Levine RJ, Karumanchi SA. Preeclampsia, a disease of the maternal endothelium: The role of antiangiogenic factors and implications for later cardiovascular disease. *Circulation*. 2011;123(24):2856-2869.

58. George EM, Arany M, Cockrell K, Storm M V., Stec DE, Granger JP. Induction of heme oxygenase-1 attenuates sFlt-1-induced hypertension in pregnant rats. *American Journal of Physiology -Regulatory, Integrative and Comparative Physiology*. 2011;301(5):R1495-R1500.

59. Cudmore M, Ahmad S, Al-Ani B, et al. Negative regulation of soluble Flt-1 and soluble endoglin release by heme oxygenase-1. *Circulation*. 2007;115(13):1789-1797.

60. Levine RJ, Maynard SE, Qian C, et al. Circulating angiogenic factors and the risk of preeclampsia. *The New England journal of medicine*. 2004;350(7):672-683.

61. Coolman M, Timmermans S, de Groot CJM, et al. Angiogenic and Fibrinolytic Factors in Blood During the First Half of Pregnancy and Adverse Pregnancy Outcomes. *Obstetrics & Gynecology*. 2012;119(6):1190-1200.

62. Lisonkova S, Joseph KS. Incidence of preeclampsia: risk factors and outcomes associated with early- versus late-onset disease. *American journal of obstetrics and gynecology*. 2013;209(6):544.e1-544.e12.

63. Valensise H, Vasapollo B, Gagliardi G, Novelli GP. Early and Late preeclampsia: Two different maternal hemodynamic states in the latent phase of the disease. *Hypertension*. 2008;52(5):873-880.

64. Nishizawa H, Pryor-Koishi K, Kato T, Kowa H, Kurahashi H, Udagawa Y. Microarray Analysis of Differentially Expressed Fetal Genes in Placental Tissue Derived from Early and Late Onset Severe Pre-eclampsia. *Placenta*. 2007;28(5-6):487-497.

65. Raymond D, Peterson E. A Critical Review of Early-Onset and Late-Onset Preeclampsia. *Obstetrical & Gynecological Survey*. 2011;66(8):497-506.

66. von Dadelszen P, Magee L, Roberts JM. Subclassification of preeclampsia. *Hypertension in pregnancy : official journal of the International Society for the Study of Hypertension in Pregnancy*. 2003;22(2):143-148.

67. Trogstad L, Magnus P, Stoltenberg C. Pre-eclampsia: Risk factors and causal models. *Best Practice and Research: Clinical Obstetrics and Gynaecology*. 2011;25(3):329-342.

68. Hernández-Díaz S, Toh S, Cnattingius S. Risk of pre-eclampsia in first and subsequent pregnancies: prospective cohort study. *BMJ (Clinical research ed)*. 2009;338:b2255.

69. Saftlas AF, Olson DR, Franks AL, Atrash HK, Pokras R. Epidemiology of preeclampsia and eclampsia in the United States, 1979-1986. *American journal of obstetrics and gynecology*. 1990;163(2):460-465.

70. Cleary-Goldman J, Malone FD, Vidaver J, et al. Impact of maternal age on obstetric outcome. *Obstetrics and gynecology*. 2005;105(5 Pt 1):983-990.

71. Conde-Agudelo A, Althabe F, Belizán JM, Kafury-Goeta AC. Cigarette smoking during pregnancy and risk of preeclampsia: A systematic review. In: *American Journal of Obstetrics and Gynecology*.Vol 181.; 1999:1026-1035.

72. Lindqvist PG, Marsál K. Moderate smoking during pregnancy is associated with a reduced risk of preeclampsia. *Acta obstetricia et gynecologica Scandinavica*. 1999;78(8):693-697.

73. Zhang J, Klebanoff M, Levine RJ, Puri M, Moyer P. The puzzling association between smoking and hypertension during pregnancy. *American journal of obstetrics and gynecology*. 1999;181(6):1407-1413.

74. England LJ, Levine RJ, Mills JL, Klebanoff MA, Yu KF, Cnattingius S. Adverse pregnancy outcomes in snuff users. *American Journal of Obstetrics and Gynecology*. 2003;189(4):939-943.

75. Engel SM, Scher E, Wallenstein S, et al. Maternal active and passive smoking and hypertensive disorders of pregnancy: risk with trimester-specific exposures. *Epidemiology (Cambridge, Mass)*. 2013;24(3):379-386.

76. Karumanchi SA, Levine RJ. How does smoking reduce the risk of preeclampsia? *Hypertension*. 2010;55(5):1100-1101.

77. Bainbridge S, Sidle EH, Smith GN. Direct placental effects of cigarette smoke protect women from pre-eclampsia: The specific roles of carbon monoxide and antioxidant systems in the placenta. *Medical Hypotheses*. 2005;64(1):17-27.

78. Kreiser D, Baum M, Seidman DS, et al. End tidal carbon monoxide levels are lower in women with gestational hypertension and pre-eclampsia. *Journal of perinatology : official journal of the California Perinatal Association*. 2004;24(4):213-217.

79. Van de Voorde J, Vanderstichele H, Leusen I. Release of endothelium-derived relaxing factor from human umbilical vessels. *Circulation research*. 1987;60(4):517-522.

80. Ahmed A, Rahman M, Zhang X, et al. Induction of placental heme oxygenase-1 is protective against TNFalpha-induced cytotoxicity and promotes vessel relaxation. *Molecular medicine (Cambridge, Mass)*. 2000;6(5):391-409.

81. Seligman SP, Buyon JP, Clancy RM, Young BK, Abramson SB. The role of nitric oxide in the pathogenesis of preeclampsia. *American journal of obstetrics and gynecology*. 1994;171(4):944-948.

82. Amaral LM, Palei ACT, Sandrim VC, et al. Maternal iNOS genetic polymorphisms and hypertensive disorders of pregnancy. *Journal of Human Hypertension*. 2012;26(9):547-552.

83. Sandrim VC, Palei AC, Cavalli RC, et al. eNOS haplotypes associated with gestational hypertension or preeclampsia. *Pharmacogenomics*. 2008;9(10):1467-1473.

84. Sandrim VC, Palei ACT, Metzger IF, Gomes V a, Cavalli RC, Tanus-Santos JE. Nitric oxide formation is inversely related to serum levels of antiangiogenic factors soluble fms-like tyrosine kinase-1 and soluble endogline in preeclampsia. *Hypertension*. 2008;52(2):402-407.

85. Sandrim VC, Montenegro MF, Palei ACT, et al. Increased circulating cell-free hemoglobin

levels reduce nitric oxide bioavailability in preeclampsia. *Free Radical Biology and Medicine*. 2010;49(3):493-500.

86. Sandrim VC, Palei ACT, Metzger IF, Cavalli RC, Duarte G, Tanus-Santos JE. Interethnic differences in ADMA concentrations and negative association with nitric oxide formation in preeclampsia. *Clinica chimica acta; international journal of clinical chemistry*. 2010;411(19-20):1457-1460.

87. Neuzil J, Stocker R. Bilirubin attenuates radical-mediated damage to serum albumin. *FEBS Letters*. 1993;331(3):281-284.

88. Neuzil J, Stocker R. Free and albumin-bound bilirubin are efficient co-antioxidants for ??-tocopherol, inhibiting plasma and low density lipoprotein lipid peroxidation. *Journal of Biological Chemistry*. 1994;269(24):16712-16719.

89. Durante W, Kroll MH, Christodoulides N, Peyton KJ, Schafer AI. Nitric oxide induces heme oxygenase-1 gene expression and carbon monoxide production in vascular smooth muscle cells. *Circulation research*. 1997;80(4):557-564.

90. George EM, Cockrell K, Adair TH, Granger JP. Regulation of sFlt-1 and VEGF secretion by adenosine under hypoxic conditions in rat placental villous explants. *American journal of physiology Regulatory, integrative and comparative physiology*. 2010;299(6):R1629-R1633.

91. Jeyabalan A, Powers RW, Durica AR, Harger GF, Roberts JM, Ness RB. Cigarette smoke exposure and angiogenic factors in pregnancy and preeclampsia. *American journal of hypertension*. 2008;21:943-947.

92. Hoyt JC, Robbins RA, Habib M, et al. Cigarette smoke decreases inducible nitric oxide synthase in lung epithelial cells. *Experimental lung research*. 2003;29(1):17-28.

93. Su Y, Han W, Giraldo C, De Li Y, Block ER. Effect of cigarette smoke extract on nitric oxide synthase in pulmonary artery endothelial cells. *American journal of respiratory cell and molecular biology*. 1998;19(5):819-825.

94. Mimura K, Tomimatsu T, Sharentuya N, et al. Nicotine restores endothelial dysfunction caused by excess sFlt1 and sEng in an in vitro model of preeclamptic vascular endothelium: a possible therapeutic role of nicotinic acetylcholine receptor (nAChR) agonists for preeclampsia. *American Journal of Obstetrics and Gynecology*. 2010;202(5):464.e1-464.e6.

95. George EM, Arany M, Cockrell K, Storm M V., Stec DE, Granger JP. Induction of heme oxygenase-1 attenuates sFlt-1-induced hypertension in pregnant rats. *AJP: Regulatory, Integrative and Comparative Physiology*. 2011;301(5):R1495-R1500.

96. Arenas-Huertero F, Apátiga-Vega E, Miguel-Pérez G, Villeda-Cuevas D, Trillo-Tinoco J. Molecular Markers Associated with the Biological Response to Aromatic Hydrocarbons from Urban Air in Humans. In: *Air Pollution - New Developments*. InTechOpen.

97. Raijmakers MTM, Peters WHM, Steegers EAP, Poston L. Amino Thiols, Detoxification and Oxidative Stress in Pre-Eclampsia and Other Disorders of Pregnancy. *Current*

Pharmaceutical Design. 2005;11:711-734.

98. Akkerman J. Nitric oxide stimulates guanylate cyclase (Pathway). *REACTOME*. 2009.

99. Kim YJ, Park HS, Lee HY, et al. Reduced L-arginine level and decreased placental eNOS activity in preeclampsia. *Placenta*. 2006;27(4-5):438-444.

100. NCBI. Heme oxygenase decycling 1, alpha subunit.

101. Zhao H, Azuma J, Kalish F, Wong RJ, Stevenson DK. Maternal Heme Oxygenase 1 Regulates Placental Vasculature Development via Angiogenic Factors in Mice. *Biology of Reproduction*. 2011;85(5):1005-1012.

102. NCBI. Hypoxia inducible factor 1. 2011.

103. Semenza GL. Hydroxylation of HIF-1: oxygen sensing at the molecular level. *Physiology (Bethesda, Md)*. 2004;19(4):176-182.

104. Brahimi-Horn MC, Pouysségur J. Oxygen, a source of life and stress. *FEBS Letters*. 2007;581(19):3582-3591.

105. Rajakumar A, Chu T, Handley DE, et al. Maternal gene expression profiling during pregnancy and preeclampsia in human peripheral blood mononuclear cells. *Placenta*. 2011;32(1):70-78.

106. Ashur-Fabian O, Yerushalmi GM, Mazaki-Tovi S, et al. Cell free expression of hif1?? and p21 in maternal peripheral blood as a marker for preeclampsia and fetal growth restriction. *PLoS ONE*. 2012;7(5):e37273.

107. Kawashima A, Koide K, Ventura W, et al. Effects of maternal smoking on the placental expression of genes related to angiogenesis and apoptosis during the first trimester. *PloS one*. 2014;9(8):e106140.

108. Ingenuity. Ingenuity Target Explorer: Xenobiotic Metabolism and Aryl Hydrocarbon Receptor Signaling Pathways. 2016.

109. Stevens EA, Mezrich JD, Bradfield CA. The aryl hydrocarbon receptor: A perspective on potential roles in the immune system. *Immunology*. 2009;127(3):299-311.

110. Dolwick KM, Schmidt J V, Carver LA, Swanson HI, Bradfield CA. Cloning and Expression of a Human Ah Receptor cDNA. *Molecular pharmacology*. 1993;44:911-917.

111. Wang K, Zhou Q, He Q, Tong G, Zhao Z, Duan T. The possible role of AhR in the protective effects of cigarette smoke on preeclampsia. *Medical Hypotheses*. 2011;77(5):872-874.

112. Ingenuity. Ingenuity Target Explorer: Nicotine Degradation II and III. 2016.

113. Pereira G, Haggar F, Shand AW, Bower C, Cook A, Nassar N. Association between pre-eclampsia and locally derived traffic-related air pollution: a retrospective cohort study. *Journal of Epidemiology & Community Health*. 2012;67(2):147-152.

114. Lee PC, Roberts JM, Catov JM, Talbott EO, Ritz B. First trimester exposure to ambient air pollution, pregnancy complications and adverse birth outcomes in Allegheny County, PA. *Maternal and Child Health Journal*. 2013;17(3):545-555.

115. Zhai D, Guo Y, Smith G, Krewski D, Walker M, Wen SW. Maternal exposure to moderate ambient carbon monoxide is associated with decreased risk of preeclampsia. *American Journal of Obstetrics and Gynecology*. 2012;207(1):57.e1-9.

116. Vinikoor-Imler LC, Gray SC, Edwards SE, Miranda ML. The effects of exposure to particulate matter and neighbourhood deprivation on gestational hypertension. *Paediatric and Perinatal Epidemiology*. 2012;26(2):91-100.

117. Rudra CB, Williams M, Sheppard L, Koenig JQ, Schiff M a. Ambient carbon monoxide and fine particulate matter in relation to preeclampsia and preterm delivery in western Washington state. *Environmental Health Perspectives*. 2011;119(6):889-892.

118. Wu J, Wilhelm M, Chung J, Ritz B. Comparing exposure assessment methods for traffic-related air pollution in an adverse pregnancy outcome study. *Environmental research*. 2011;111(5):685-692.

119. Wu J, Ren C, Delfino RJ, Chung J, Wilhelm M, Ritz B. Association between local traffic-generated air pollution and preeclampsia and preterm delivery in the South Coast Air Basin of california. *Environmental Health Perspectives*. 2009;117(11):1773-1779.

120. van den Hooven EH, Jaddoe VW V, de Kluizenaar Y, et al. Residential traffic exposure and pregnancy-related outcomes: a prospective birth cohort study. *Environmental health : a global access science source*. 2009;8:59.

121. Savitz DA, Elston B, Bobb JF, Johnson S, Mcalexander T. Ambient fine particulate matter, nitrogen dioxide, and hypertensive disorders of pregnancy in New York City. *Epidemiology*. 2015;26(5):748-757.

122. Parazzini F, Ricci E, Chatenoud L, et al. Maternal and paternal smoking and pregnancy-induced hypertension. *European Journal of Obstetrics Gynecology and Reproductive Biology*. 2003;109(2):141-144.

123. Chappell S, Morgan L. Searching for genetic clues to the causes of pre-eclampsia. *Clinical science (London, England : 1979)*. 2006;110(4):443-458.

124. Goddard KB, Tromp G, Romero R, et al. Candidate-gene association study of mothers with pre-eclampsia, and their infants, analyzing 775 SNPs in 190 genes. *Human Heredity*. 2007;63(1):1-16.

125. Cnattingius S, Reilly M, Pawitan Y, Lichtenstein P. Maternal and fetal genetic factors account for most of familial aggregation of preeclampsia: A population-based swedish cohort study. *American Journal of Medical Genetics*. 2004;130 A(4):365-371.

126. Ros HS, Lichtenstein P, Lipworth L, Cnattingius S. Genetic effects on the liability of developing pre-eclampsia and gestational hypertension. *American Journal of Medical Genetics*. 2000;91(4):256-260.

127. Lie RT, Rasmussen S, Brunborg H, Gjessing HK, Lie-Nielsen E, Irgens LM. Fetal and maternal contributions to risk of pre-eclampsia: population based study. *BMJ (Clinical research ed)*. 1998;316(7141):1343-1347.

128. Esplin MS, Fausett MB, Fraser A, et al. Paternal and maternal components of the predisposition to preeclampsia. *The New England journal of medicine*. 2001;344(12):867-872.

129. Takahashi K, Kobayashi T, Kanayama N. p57(Kip2) regulates the proper development of labyrinthine and spongiotrophoblasts. *Molecular human reproduction*. 2000;6(11):1019-1025.

130. Tuteja G, Cheng E, Papadakis H, Bejerano G. PESNPdb: A comprehensive database of SNPs studied in association with pre-eclampsia. *Placenta*. 2012;33(12):1055-1057.

131. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. *Nature genetics*. 2004;36(5):431-432.

132. Arngrímsson R, Siguröardóttir S, Frigge ML, et al. A genome-wide scan reveals a maternal susceptibility locus for pre-eclampsia on chromosome 2p13. *Human Molecular Genetics*. 1999;8(9):1799-1805.

133. Moses EK, Lade J, Guo G, et al. A genome scan in families from Australia and New Zealand confirms the presence of a maternal susceptibility locus for pre-eclampsia, on chromosome 2. *American journal of human genetics*. 2000;67(6):1581-1585.

134. Laivuori H, Lahermo P, Ollikainen V, et al. Susceptibility loci for preeclampsia on chromosomes 2p25 and 9p13 in Finnish families. *American journal of human genetics*. 2003;72(1):168-177.

135. Roten LT, Johnson MP, Forsmo S, et al. Association between the candidate susceptibility gene ACVR2A on chromosome 2q22 and pre-eclampsia in a large Norwegian population-based study (the HUNT study). *European journal of human genetics : EJHG*. 2009;17(2):250-257.

136. Johnson MP, Brennecke SP, East CE, et al. Genome-wide association scan identifies a risk locus for preeclampsia on 2q14, near the inhibin, beta B gene. Grant SFA, ed. *PLoS ONE*. 2012;7(3):e33666.

137. Lachmeijer AM, Crusius JB, Pals G, Dekker GA, Arngrímsson R, ten Kate LP. Polymorphisms in the tumor necrosis factor and lymphotoxin-alpha gene region and preeclampsia. *Obstetrics and gynecology*. 2001;98(4):612-619.

138. Haggerty CL, Ferrell RE, Hubel C a, Markovic N, Harger G, Ness RB. Association between allelic variants in cytokine genes and preeclampsia. *American Journal of Obstetrics and Gynecology*. 2005;193(1):209-215.

139. Mirahmadian M, Kalantar F, Heidari G, Safdarian L, Mansouri R, Amirzargar AA. Association of tumor necrosis factor-alpha and interleukin-10 gene polymorphisms in Iranian patients with pre-eclampsia. *American journal of reproductive immunology : AJRI : official journal of the American Society for the Immunology of Reproduction and the*

International Coordination Committee for Immunology of Reproduction. 2008;60(2):179-185.

140. Lévesque S, Moutquin JM, Lindsay C, Roy MC, Rousseau F. Implication of an AGT Haplotype in a Multigene Association Study with Pregnancy Hypertension. *Hypertension*. 2004;43(1):71-78.

141. Johnson MP, Roten LT, Dyer TD, et al. The ERAP2 gene is associated with preeclampsia in Australian and Norwegian populations. *Human Genetics*. 2009;126(5):655-666.

142. Buurma AJ, Turner RJ, Driessen JHM, et al. Genetic variants in pre-eclampsia: A meta-analysis. *Human Reproduction Update*. 2013;19(3):289-303.

143. Agorastos T, Karavida A, Lambropoulos A, et al. Factor V Leiden and prothrombin G20210A mutations in pregnancies with adverse outcome. *The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians*. 2002;12(4):267-273.

144. Gerhardt A, Goecke TW, Beckmann MW, et al. The G20210A prothrombin-gene mutation and the plasminogen activator inhibitor (PAI-1) 5G/5G genotype are associated with early onset of severe preeclampsia. *Journal of Thrombosis and Haemostasis*. 2005;3(4):686-691.

145. Dusse LMS, Carvalho MDG, Bragança WF, et al. Inherited thrombophilias and pre-eclampsia in Brazilian women. *European Journal of Obstetrics Gynecology and Reproductive Biology*. 2007;134(1):20-23.

146. Dizon-Townson DS, Kinney S, Branch DW, Ward K. The factor V Leiden mutation is not a common cause of recurrent miscarriage. *Journal of reproductive immunology*. 1997;34(3):217-223.

147. Livingston JC, Barton JR, Park V, Haddad B, Phillips O, Sibai BM. Maternal and fetal inherited thrombophilias are not related to the development of severe preeclampsia. *American Journal of Obstetrics and Gynecology*. 2001;185(1):153-157.

148. Heiskanen JTM, Pirskanen MM, Hiltunen MJ, Mannermaa AJ, Punnonen KRA, Heinonen ST. Insertion-deletion polymorphism in the gene for angiotensin-converting enzyme is associated with obstetric cholestasis but not with preeclampsia. *American Journal of Obstetrics and Gynecology*. 2001;185(3):600-603.

149. Kobashi G, Yamada H, Ohta K, Kato E, Ebina Y, Fujimoto S. Endothelial nitric oxide synthase gene (NOS3) variant and hypertension in pregnancy. *American journal of medical genetics*. 2001;103(3):241-244.

150. Higgins JR, Kaiser T, Moses EK, North R, Brennecke SP. Prothrombin G20210A mutation: is it associated with pre-eclampsia? *Gynecologic and obstetric investigation*. 2000;50(4):254-257.

151. Hubel CA, Roberts JM, Ferrell RE. Association of pre-eclampsia with common coding sequence variations in the lipoprotein lipase gene. *Clinical genetics*. 1999;56(4):289-296.

152. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Lippincott Williams & Wilkins; 2008.

153. Aggarwal PK, Jain V, Jha V. Endothelial nitric oxide synthase, angiotensin-converting enzyme and angiotensinogen gene polymorphisms in hypertensive disorders of pregnancy. *Hypertension research : official journal of the Japanese Society of Hypertension*. 2010;33(5):473-477.

154. Landau R, Xie HG, Dishy V, Wood AJJ, Stein CM, Smiley RM. No association of the Asp298 variant of the endothelial nitric oxide synthase gene with preeclampsia. *American Journal of Hypertension*. 2004;17(5):391-394.

155. Serrano NC, Casas JP, Díaz L, et al. Endothelial NO synthase genotype and risk of preeclampsia: A multicenter case-control study. *Hypertension*. 2004;44(5):702-707.

156. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet*. 2003;361(9357):598-604.

157. Zavattari P, Deidda E, Whalen M, et al. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Human molecular genetics*. 2000;9(20):2947-2957.

158. Zhao L, Triche EW, Walsh KM, et al. Genome-wide association study identifies a maternal copy-number deletion in PSG11 enriched among preeclampsia patients. *BMC Pregnancy and Childbirth*. 2012;12(1):61.

159. Zhao L, Bracken MB, DeWan AT. Genome-Wide Association Study of Pre-Eclampsia Detects Novel Maternal Single Nucleotide Polymorphisms and Copy-Number Variants in Subsets of the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study Cohort. *Annals of Human Genetics*. 2013;77(4):277-287.

160. Coolman M, de Groot CJM, Jaddoe VW, Hofman A, Raat H, Steegers E a P. Medical record validation of maternally reported history of preeclampsia. *Journal of Clinical Epidemiology*. 2010;63(8):932-937.

161. Smith CJ, Saftlas AF, Spracklen CN, et al. Genetic Risk Score for Essential Hypertension and Risk of Preeclampsia. *American Journal of Hypertension*. 2015;29(January):1-8.

162. Charles IG, Scorer C, Moro M, et al. Expression of human nitric oxide synthase isozymes. *Methods in enzymology*. 1996;268(1992):449-460.

163. Veldman B, Spiering W, Doevendans P, et al. The Glu298Asp polymorphism of the NOS 3 gene as a determinant of the baseline production of nitric oxide. *Journal of hypertension*. 2002;20(10):2023-2027.

164. Qi H-P, Fraser WD, Luo Z-C, Julien P, Audibert F, Wei S-Q. Endothelial nitric oxide synthase gene polymorphisms and risk of preeclampsia. *American journal of perinatology*. 2013;30(10):795-804.

165. Durante W, Christodoulides N, Cheng K, Peyton KJ, Sunahara RK, Schafer a I. cAMP

induces heme oxygenase-1 gene expression and carbon monoxide production in vascular smooth muscle. *The American journal of physiology*. 1997;273(1 Pt 2):H317-H323.

166. Kuhlencordt PJ, Gyurko R, Han F, et al. Accelerated atherosclerosis, aortic aneurysm formation, and ischemic heart disease in apolipoprotein E/endothelial nitric oxide synthase double-knockout mice. *Circulation*. 2001;104(4):448-454.

167. Manso H, Krug T, Sobral J, et al. Variants within the nitric oxide synthase 1 gene are associated with stroke susceptibility. *Atherosclerosis*. 2012;220(2):443-448.

168. Lin R, Fu W, Zhou W, et al. Association of heme oxygenase-1 gene polymorphisms with essential hypertension and blood pressure in the Chinese Han population. *Genetic testing and molecular biomarkers*. 2011;15(1-2):23-28.

169. Ellinor PT, Lunetta KL, Albert CM, et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nature Genetics*. 2012;44(6):670-675.

170. Holm H, Gudbjartsson DF, Arnar DO, et al. Several common variants modulate heart rate, PR interval and QRS duration. *Nature genetics*. 2010;42(2):117-122.

171. Aouizerat BE, Vittinghoff E, Musone SL, et al. GWAS for discovery and replication of genetic loci associated with sudden cardiac arrest in patients with coronary artery disease. *BMC cardiovascular disorders*. 2011;11(1):29.

172. Lu X, Wang L, Chen S, et al. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nature Genetics*. 2012;44(8):890-894.

173. Ehret GB, Munroe PB, Rice KM, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478(7367):103-109.

174. Wain L V, Verwoert GC, O'Reilly PF, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nature Genetics*. 2011;43(10):1005-1011.

175. Wojczynski MK, Li M, Bielak LF, et al. Genetics of coronary artery calcification among African Americans, a meta-analysis. *BMC medical genetics*. 2013;14(1):75.

176. Verweij N, Mahmud H, Mateo Leach I, et al. Genome-wide association study on plasma levels of midregional-proadrenomedullin and C-terminal-pro-endothelin-1. *Hypertension*. 2013;61(3):602-608.

177. Ganesh SK, Zakai NA, van Rooij FJA, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nature genetics*. 2009;41(11):1191-1198.

178. Arngrímsson R, Hayward C, Nadaud S, et al. Evidence for a familial pregnancy-induced hypertension locus in the eNOS-gene region. *American journal of human genetics*. 1997;61(2):354-362.

179. Chen L-K, Huang C-H, Yeh H-M, et al. Polymorphisms in the endothelial nitric oxide

synthase gene may be protective against preeclampsia in a Chinese population. *Reproductive sciences (Thousand Oaks, Calif)*. 2007;14(2):175-181.

180. Díaz-Olguín L, Coral-Vázquez RM, Canto-Cetina T, et al. Endothelial nitric oxide synthase haplotypes are associated with preeclampsia in Maya mestizo women. *Disease Markers*. 2011;31(2):83-89.

181. Fatini C, Sticchi E, Gensini F, et al. Endothelial nitric oxide synthase gene influences the risk of pre-eclampsia, the recurrence of negative pregnancy events, and the maternal-fetal flow. *Journal of hypertension*. 2006;24(9):1823-1829.

182. Nishizawa H, Pryor-Koishi K, Suzuki M, et al. Analysis of nitric oxide metabolism as a placental or maternal factor underlying the etiology of pre-eclampsia. *Gynecologic and Obstetric Investigation*. 2009;68(4):239-247.

183. Pappa KI, Roubelakis M, Vlachos G, et al. Variable effects of maternal and paternal-fetal contribution to the risk for preeclampsia combining GSTP1, eNOS, and LPL gene polymorphisms. *The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians*. 2011;24(4):628-635.

184. Sharma D, Trivedi SS, Bhattacharjee J. Oxidative stress and eNOS (Glu298Asp) gene polymorphism in preeclampsia in Indian population. *Molecular and Cellular Biochemistry*. 2011;353(1-2):189-193.

185. Zdoukopoulos N, Doxani C, Messinis IE, Stefanidis I, Zintzaras E. Polymorphisms of the endothelial nitric oxide synthase (NOS3) gene in preeclampsia: a candidate-gene association study. *BMC Pregnancy and Childbirth*. 2011;11(1):89.

186. Goddard KB, Tromp G, Romero R, et al. Candidate-Gene Association Study of Mothers with Pre-Eclampsia, and Their Infants, Analyzing 775 SNPs in 190 Genes. *Human Heredity*. 2007;63(1):1-16.

187. Lachmeijer AMA, Nosti-Escanilla MP, Bastiaans EB, et al. Linkage and association studies of IL1B and IL1RN gene polymorphisms in preeclampsia. *Hypertension in pregnancy : official journal of the International Society for the Study of Hypertension in Pregnancy*. 2002;21(1):23-38.

188. The GOPEC Consortium. Disentangling fetal and maternal susceptibility for pre-eclampsia: a British multicenter candidate-gene study. *American journal of human genetics*. 2005;77(1):127-131.

189. Shimodaira M, Nakayama T, Sato I, et al. Estrogen synthesis genes CYP19A1, HSD3B1, and HSD3B2 in hypertensive disorders of pregnancy. *Endocrine*. 2012;42(3):700-707.

190. Kim YJ, Park HS, Park MH, Suh SH, Pang MG. Oxidative stress-related gene polymorphism and the risk of preeclampsia. *European Journal of Obstetrics Gynecology and Reproductive Biology*. 2005;119(1):42-46.

191. Cummings AM, Kavlock RJ. Gene-environment interactions: a review of effects on

reproduction and development. *Critical reviews in toxicology*. 2004;34(6):461-485.

192. Barrett JH, Sheehan NA, Cox A, Worthington J, Cannings C, Teare MD. Family based studies and genetic epidemiology: Theory and practice. *Human Heredity*. 2007;64(2):146-148.

193. Tiret L, Abel L, Rakotovao R. Effect of ignoring genotype-environment interaction on segregation analysis of quantitative traits. In: *Genetic Epidemiology*.Vol 10.; 1993:581-586.

194. Hill LD, Hilliard DD, York TP, et al. Fetal ERAP2 variation is associated with preeclampsia in African Americans in a case-control study. *BMC medical genetics*. 2011;12(1):64.

195. Petry CJ, Sanz Marcos N, Pimentel G, et al. Associations Between Fetal Imprinted Genes and Maternal Blood Pressure in Pregnancy. *Hypertension*. 2016:HYPERTENSIONAHA.116.08261.

196. Wan J, Zhao H, Li T, Li C, Wang X, Chen Z-J. The common variant rs11646213 is associated with preeclampsia in Han Chinese women. *PloS one*. 2013;8(8):e71202.

197. Wan J-P, Wang H, Li C-Z, et al. The Common Single-Nucleotide Polymorphism rs2681472 Is Associated With Early-Onset Preeclampsia in Northern Han Chinese Women. *Reproductive sciences (Thousand Oaks, Calif)*. 2014;21(11):1423-1427.

198. Bhatnagar S, Bhattacharjee J, Vaid M, Madan T, Trivedi SS, Sarma PU. Inducible nitric oxide synthase (iNOS) gene polymorphism in pre-eclampsia: a pilot study in North India. *The Australian & New Zealand journal of obstetrics & gynaecology*. 2007;47(6):477-482.

199. Srinivas SK, Morrison AC, Andrela CM, Elovitz M a. Allelic variations in angiogenic pathway genes are associated with preeclampsia. *American Journal of Obstetrics and Gynecology*. 2010;202(5):445.e1-11.

200. Herse F, Dechend R, Harsem NK, et al. Dysregulation of the circulating and tissue-based renin-angiotensin system in preeclampsia. In: *Hypertension*.Vol 49.; 2007:604-611.

201. Han JY, Kim YS, Cho GJ, et al. Altered gene expression of caspase-10, death receptor-3 and IGFBP-3 in preeclamptic placentas. *Molecules and cells*. 2006;22(2):168-174.

202. Zhou R, Zhu Q, Wang Y, Ren Y, Zhang L, Zhou Y. Genomewide oligonucleotide microarray analysis on placentae of pre-eclamptic pregnancies. *Gynecologic and Obstetric Investigation*. 2006;62(2):108-114.

203. Hansson SR, Chen Y, Brodszki J, et al. Gene expression profiling of human placentas from preeclamptic and normotensive pregnancies. *Molecular Human Reproduction*. 2006;12(3):169-179.

204. Gack S, Marmé A, Marmé F, et al. Preeclampsia: Increased expression of soluble ADAM 12. *Journal of Molecular Medicine*. 2005;83(11):887-896.

205. Soleymanlou N, Jurisica I, Nevo O, et al. Molecular evidence of placental hypoxia in preeclampsia. *The Journal of clinical endocrinology and metabolism*. 2005;90(7):4299-

4308.

206. Heikkilä A, Tuomisto T, Häkkinen SK, Keski-Nisula L, Heinonen S, Ylä-Herttuala S. Tumor suppressor and growth regulatory genes are overexpressed in severe early-onset preeclampsia - An array study on case-specific human preeclamptic placental tissue. *Acta Obstetricia et Gynecologica Scandinavica*. 2005;84(7):679-689.

207. Pang ZJ, Xing FQ. Comparative profiling of metabolism-related gene expression in pre-eclamptic and normal pregnancies. *Archives of Gynecology and Obstetrics*. 2004;269(2):91-95.

208. Tsoi SCM, Cale JM, Bird IM, Kay HH. CDNA microarray analysis of gene expression profiles in human placenta: Up-regulation of the transcript encoding muscle subunit of glycogen phosphorylase in preeclampsia. *Journal of the Society for Gynecologic Investigation*. 2003;10(8):496-502.

209. Pang ZJ, Xing FQ. Expression profile of trophoblast invasion-associated genes in the pre-eclamptic placenta. *British Journal of Biomedical Science*. 2003;60(2):97-101.

210. Reimer T, Koczan D, Gerber B, Richter D, Thiesen HJ, Friese K. Microarray analysis of differentially expressed genes in placental tissue of pre-eclampsia: up-regulation of obesity-related genes. *Molecular human reproduction*. 2002;8(7):674-680.

211. Harlow FH, Brown MA. The diversity of diagnoses of preeclampsia. *Hypertension in pregnancy : official journal of the International Society for the Study of Hypertension in Pregnancy*. 2001;20(1):57-67.

212. Weinberg CR, Shi M. The genetics of preterm birth: Using what we know to design better association studies. *American Journal of Epidemiology*. 2009;170(11):1373-1381.

213. Shi M, Umbach DM, Vermeulen SH, Weinberg CR. Making the most of case-mother/control-mother studies. *American Journal of Epidemiology*. 2008;168(5):541-547.

214. Phillips G, Kelly A, Pitt JA, Sanderson R, Taylor E. The recovery of a very shallow eutrophic lake, 20 years after the control of effluent derived phosphorus. *Freshwater Biology*. 2005;50(10):1628-1638.

215. Tsai HJ, Liu X, Mestan K, et al. Maternal cigarette smoking, metabolic gene polymorphisms, and preterm delivery: New insights on G × E interactions and pathogenic pathways. *Human Genetics*. 2008;123(4):359-369.

216. Hutter CM, Mechanic LE, Chatterjee N, et al. Gene-environment interactions in cancer epidemiology: A national cancer institute think tank report. *Genetic Epidemiology*. 2013;37(7):643-657.

217. Umbach DM, Weinberg CR. The use of case-parent triads to study joint effects of genotype and exposure. *American journal of human genetics*. 2000;66(1):251-261.

218. Kistner EO, Shi M, Weinberg CR. Using cases and parents to study multiplicative gene-by-environment interaction. *American Journal of Epidemiology*. 2009;170(3):393-400.

219. Norwegian Institute of Public Health. *Norwegian Mother and Child Cohort Study: Revised Protocol, End of Enrollment - Protocol II*.; 2012.

220. Magnus P, Irgens LM, Haug K, et al. Cohort profile: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology*. 2006;35(5):1146-1150.

221. Rønningen KS, Paltiel L, Meltzer HM, et al. The biobank of the Norwegian mother and child cohort Study: A resource for the next 100 years. *European Journal of Epidemiology*. 2006;21(8):619-625.

222. Magnus P, Birke C, Vejrup K, et al. Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology*. 2016;45(2):382-388.

223. Klungsøyr K, Harmon QE, Skard LB, et al. Validity of Pre-Eclampsia Registration in the Medical Birth Registry of Norway for Women Participating in the Norwegian Mother and Child Cohort Study, 1999-2010. *Paediatric and Perinatal Epidemiology*. September 2014:362-371.

224. Beadchip IC-. Infinium CoreExome-24 v1.1 Data Sheet. 2011;(June):2011-2012.

225. Grove ML, Yu B, Cochran BJ, et al. Best Practices and Joint Calling of the HumanExome BeadChip: The CHARGE Consortium. *PLoS ONE*. 2013;8(7):e68095.

226. Devlin B, Roeder K. Genomic Control for Association Studies. *Biometrics*. 1999;55(4):997-1004.

227. Kvalvik LG, Nilsen RM, Skjærven R, et al. Self-reported smoking status and plasma cotinine concentrations among pregnant women in the Norwegian Mother and Child Cohort Study. *Pediatric Research*. 2012;72(1):101-107.

228. Wang Y, Zhao S. *Vascular Biology of the Placenta*. (Granger DN, Granger JP, eds.). Morgan and Claypool Life Sciences; 2010.

229. Wilson ML, Goodwin TM, Pan VL, Ingles SA. Molecular epidemiology of preeclampsia. *Obstetrical & gynecological survey*. 2003;58(1):39-66.

230. Saftlas AF, Beydoun H, Triche E. Immunogenetic determinants of preeclampsia and related pregnancy disorders: a systematic review. *Obstetrics and gynecology*. 2005;106(1):162-172.

231. Chen J, Zheng H, Wilson ML. Likelihood ratio tests for maternal and fetal genetic effects on obstetric complications. *Genetic Epidemiology*. 2009;33(6):526-538.

232. Ainsworth HF, Unwin J, Jamison DL, Cordell HJ. Investigation of maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting), using mothers and their offspring. *Genetic Epidemiology*. 2011;35(1):19-45.

233. Wilcox AJ, Weinberg CR, Lie RT. Distinguishing the effects of maternal and offspring genes through studies of “case-parent triads”. *American journal of epidemiology*. 1998;148(9):893-901.

234. Weinberg CR, Wilcox AJ, Lie RT. A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *American journal of human genetics*. 1998;62(4):969-978.

235. van Den Oord EJ, Vermunt JK. Testing for linkage disequilibrium, maternal effects, and imprinting with (In)complete case-parent triads, by use of the computer program LEM. *American journal of human genetics*. 2000;66(1):335-338.

236. Vermunt JK. LEM 1.0 : A general program for the analysis of categorical data. 1997.

237. Humphreys K, Grankvist A, Leu M, et al. The genetic structure of the Swedish population. *PLoS ONE*. 2011;6(8):e22547.

238. McEvoy BP, Montgomery GW, McRae AF, et al. Geographical structure and differential natural selection among North European populations. *Genome Research*. 2009;19(5):804-814.

239. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995;57(1):289-300.

240. Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. *Statistics in Medicine*. 1983;2(1983):243-251.

241. Morgan L, McGinnis R, Steinthorsdottir V, et al. InterPregGen: Genetic studies of pre-eclampsia in three continents. *Norsk Epidemiologi*. 2014;24(1-2):141-146.

242. Genetics T, Gopec P. Babies, pre-eclamptic mothers and grandparents: a three-generation phenotyping study. *Journal of hypertension*. 2007;25(4):849-854.

243. Harmon Q, Huang L, Umbach D, Klungsoyr K, Engel S, Magnus P. Risk of Fetal Death with Preeclampsia. *Obstetrics and Gynecology*. 2015;125(3):628-635.

244. Ruano R, Fontes RS, Zugaib M. Prevention of preeclampsia with low-dose aspirin -- a systematic review and meta-analysis of the main randomized controlled trials. *Clinics (São Paulo, Brazil)*. 2005;60(5):407-414.

245. Gilbert JSJ, Ryan MJM, LaMarca BBB, et al. Pathophysiology of hypertension during preeclampsia: linking placental ischemia with endothelial dysfunction. *American journal of physiology Heart and circulatory physiology*. 2008;294(2):H541-H550.

246. Wikström A-K, Stephansson O, Cnattingius S. Tobacco use during pregnancy and preeclampsia risk: effects of cigarette smoking and snuff. *Hypertension*. 2010;55(5):1254-1259.

247. Irgens LM. The Medical Birth Registry of Norway. Epidemiological research and surveillance throughout 30 years. *Acta obstetricia et gynecologica Scandinavica*. 2000;79(6):435-439.

248. Vermeulen SH, Shi M, Weinberg CR, Umbach DM. A hybrid design: Case-parent triads

supplemented by control-mother dyads. *Genetic Epidemiology*. 2009;33(2):136-144.

249. Hanoune J, Defer N. Regulation and Role of Adenylyl Cyclase Isoforms. *Annual reviews of pharmacologic toxicology*. 2001;41:145-174.

250. Lucas KA, Pitari GM, Kazerounian S, et al. Guanylyl Cyclases and Signaling by Cyclic GMP. 2000;52(3):375-413.

251. Lugnier C. Cyclic nucleotide phosphodiesterase (PDE) superfamily : A new target for the development of specific therapeutic agents. *Pharmacology and Therapeutics*. 2006;109:366-398.

252. Ghofrani HA, Osterloh IH, Grimminger F. Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond. *Nature reviews Drug discovery*. 2006;5(8):689-702.

253. Motta C, Grosso C, Zanuzzi C, et al. Effect of Sildenafil on Pre-Eclampsia-Like Mouse Model Induced By L-Name. *Reproduction in Domestic Animals*. 2015;616:611-616.

254. Stanley JL, Sulek K, Andersson IJ, et al. Sildenafil Therapy Normalizes the Aberrant Metabolomic Profile in the Comt − / − Mouse Model of Preeclampsia / Fetal Growth Restriction. *Nature Publishing Group*. 2015;(November):1-10.

255. Academisch Medisch Centrum - Universiteit van Amsterdam. The Dutch STRIDER (Sildenafil TheRapy in Dismal Prognosis Early-onset Fetal Growth Restriction) (STRIDER). *ClinicalTrials.gov*. 2014.

256. Assiut University. Sildenafil Citrate for the Management of Asymmetrical Intrauterine Growth Restriction. *ClinicalTrials.gov*. 2016.

257. Samangaya R, Mires G, Shennan A, et al. A randomised, double-blinded, placebo-controlled study of the phosphodiesterase type 5 inhibitor sildenafil for the treatment of preeclampsia. *Hypertension in pregnancy*. 2009;28(January 2016):369-382.

258. Nicolae DL, Gamazon E, Zhang W, Duan S, Eileen Dolan M, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genetics*. 2010;6(4).

259. Shi M, Murray JC, Marazita ML, et al. the etiology of orofacial clefts. 2013;0(4):784-794.

260. Kim YJ, Williamson R, Chen K, Smith JL, Murray JC, Merrill DC. Lipoprotein lipase gene mutations and the genetic susceptibility of preeclampsia. *Hypertension*. 2001;38(5):992-996.

261. Lisonkova S, Joseph KS. Left Truncation Bias as a Potential Explanation for the Protective Effect of Smoking on Preeclampsia. *Epidemiology*. 2015;26(3):436-440.

262. Shi M, Umbach DM, Weinberg CR. Family-based Gene-by-Environment Interaction Studies: Revelations and Remedies. *Epidemiology*. 2011;22(3):400-407.

263. Gerhardt K. The Placenta: A Bridge Between Mother and Baby. *Biology of Reproduction*. 2016;95(6):114-114.

264. Sferruzzi-Perri AN, López-Tello J, Fowden AL, Constancia M. Maternal and fetal genomes interplay through phosphoinositol 3-kinase(PI3K)-p110α signaling to modify placental resource allocation. *Proceedings of the National Academy of Sciences*. 2016;113(40):201602012.

265. Marlowe JL, Fan Y, Chang X, et al. The Aryl Hydrocarbon Receptor Binds to E2F1 and Inhibits E2F1-induced Apoptosis. *Molecular biology of the cell*. 2008;19:3263-3271.

266. Taioli E. Gene – environment interaction in tobacco-related cancers. *Carcinogenesis*. 2008;29(8):1467-1474.

267. Zhan P, Wang Q, Qian Q, Wei S, Yu L. CYP1A1 MspI and exon7 gene polymorphisms and lung cancer risk : An updated meta-analysis and review. *Journal of experimental and clinical cancer research*. 2011;30:99.

268. Zhang Z, Hao K, Shi R, et al. Human Genome Epidemiology (HuGE) Review Glutathione S -Transferase M1 (GSTM1) and Glutathione S -Transferase T1 (GSTT1) Null Polymorphisms , Smoking , and Their Interaction in Oral Cancer : A HuGE Review and Meta-Analysis. *American Journal of Epidemiology*. 2011;173(8):847-857.

269. Geisler NC, Geisler SA, Olshan AF. GSTM1, GSTT1, and the Risk of Squamous Cell Carcinoma of the Head and Neck: A Mini-HuGE Review. *American Journal of Epidemiology*. 2001;154(2).

270. Raimondi S, Botteri E, Iodice S, Lowenfels AB, Maisonneuve P. Gene-smoking interaction on colorectal adenoma and cancer risk : Review and meta-analysis. *Mutation Research*. 2009;670:6-14.

271. Luef BM, Andersen LB, Renault KM, Nohr EA, Jorgensen JS, Christesen HT. Validation of hospital discharge diagnoses for hypertensive disorders of pregnancy. *Acta Obstetricia et Gynecologica Scandinavica*. 2016;95:1288-1294.

272. England LJ, Grauman A, Qian C, et al. Misclassification of maternal smoking status and its effects on an epidemiologic study of pregnancy outcomes. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco*. 2007;9(10):1005-1013.

273. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: Practical considerations. *Nature Reviews Genetics*. 2002;3(5):391-397.

274. Vefring HK, Wee L, Jugessur A, Gjessing HK, Nilsen ST, Lie RT. Maternal angiotensinogen (AGT) haplotypes, fetal renin (REN) haplotypes and risk of preeclampsia; estimation of gene-gene interaction from family-triad data. *BMC Medical Genetics.* 2010;11:90.

275. Li A, Meyre D. Challenges in reproducibility of genetic association studies: lessons learned from the obesity field. *International Journal of Obesity*. 2012;37(4):559-567.

276. Vanderweele TJ, Knol MJ. A Tutorial on Interaction. *Epidemiologic Methods*. 2014;3(1):33-72.

277. Nilsen RM, Vollset SE, Gjessing HK, et al. Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatric and Perinatal Epidemiology*. 2009;23(6):597-608.

278. Kinlaw AC, Buckley JP, Engel SM, Poole C, Brookhart MA, Keil AP. Left Truncation Bias as an Explanation for the Protective Effect of Smoking on Preeclampsia: Potential, but How Plausible? *Epidemiology*. in press.

279. Zusterzeel PLM, Nelen WLDM, Roelofs HMJ, Peters WHM, Blom HJ, Steegers EAP. Polymorphisms in biotransformation enzymes and the risk for recurrent early pregnancy loss. *Molecular Human Reproduction*. 2000;6(5):474-478.

280. Williams PJ, Broughton Pipkin F. The genetics of pre-eclampsia and other hypertensive disorders of pregnancy. *Best Practice and Research: Clinical Obstetrics and Gynaecology*. 2011;25(4):405-417.

281. Ward LD, Kellis M. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Research*. 2012;40(D1):930-934.

282. Sinsheimer JS, Elston RC, Fu WJ. Gene-gene interaction in maternal and perinatal research. *Journal of Biomedicine and Biotechnology*. 2010;2010.

283. Sinsheimer JS, Palmer CGS, Woodward JA. Detecting genotype combinations that increase risk for disease: The maternal-fetal genotype incompatibility test. *Genetic Epidemiology*. 2003;24(1):1-13.

284. Clark MM, Blangero J, Dyer TD, Sobel EM, Sinsheimer JS. The Quantitative-MFG Test: A linear mixed effect model to detect maternal-offspring gene interactions. *Ann Hum Genet*. 2016;80(1):63-80.

285. Dekker G, Robillard PY, Roberts C. The etiology of preeclampsia: The role of the father. *Journal of Reproductive Immunology*. 2011;89(2):126-132.

286. Wang X, Miller DC, Harman R, Antczak DF, Clark AG. Paternally expressed genes predominate in the placenta. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(26):10705-10710.

287. Skjaerven R, Vatten LJ, Wilcox AJ, Rønning T, Irgens LM, Lie RT. Recurrence of pre-eclampsia across generations: exploring fetal and maternal genetic components in a population based cohort. *BMJ (Clinical research ed.)*. 2005;331(September):877.

288. Macdonald-Wallis C, Lawlor DA, Heron J, Fraser A, Nelson SM, Tilling K. Relationships of risk factors for pre-eclampsia with patterns of occurrence of isolated gestational proteinuria during normal term pregnancy. *PLoS ONE*. 2011;6(7):1-10.

289. Frias A, Schabel M, Roberts V, et al. Using Dynamic Contrast Enhanced MRI to Quantitatively Characterize Maternal Vascular Organization in the Primate Placenta. *Magn Reson Med*. 2015;73(4):1570-1578.

290. Osmanski B-F, Lecarpentier E, Montaldo G, Tsatsaris V, Chavatte-Palmer P, Tanter M. Discriminative imaging of maternal and fetal blood flow within the placenta using ultrafast

ultrasound. *Scientific reports*. 2015;5:13394.

291. Manning AK, Lavalley M, Liu C, et al. Meta-analysis of Gene-Environment interaction: joint estimation of SNP and SNP×Environment regression coefficients. *Genetic Epidemiology*. 2012;35(1):11-18.

292. Brooks SA, Martin E, Smeester L, Grace MR, Boggess K, Fry RC. miRNAs as common regulators of the transforming growth factor (TGF)-β pathway in the preeclamptic placenta and cadmium-treated trophoblasts: Links between the environment, the epigenome and preeclampsia. *Food and Chemical Toxicology*. 2016;98:50-57.